Contents

Part I

Econometrics

Chapter 1

Common Tools

Seemingly Unrelated Regressions: [1]

- $y_i = X_i^T \theta_0 + u_i$, where X_i is k by 1.
- $\hat{\theta} = (\sum_{i=1}^n X_i \hat{\Psi}^{-1} X_i^T)^{-1} \sum_{i=1}^n X_i \hat{\Psi}^{-1} y_i$
- This estimator is more efficient than OLS. And it only requires the estimation of unconditional variance of error term, while WLS requires the whole distribution of error term conditional on X_i. So will also be more efficient than WLS in practice.
- Intuitively, it uses the correlation information of error terms across equations.
- Notice, more applicable when k is not very large (relative to n). Otherwise the estimation of $\hat{\Psi}$ will be too demanding.

[1] see Pinkse notes.

Chapter 2

Identification

2.1 Identification in Multi-Equilibria

- Find some feature that is common to all equilibria. For example, Berry (1992)

- Set identification. For example, Iaryczower et al. (2018)

2.2 Lewbel, 2016, JEL

Lewbel, A. (2016). The Identification Zoo-Meanings of Identification in Econometrics. Forthcoming on Journal of Economic Literature

`https://www2.bc.edu/arthur-lewbel/ident-zoo-SL-Part1.pdf`

`https://www2.bc.edu/arthur-lewbel/ident-zoo-SL-Part2.pdf`

are links for a ppt on this paper by Lewbel.

There are two kinds of identification problems. 1. One is to identify the treatment effect, a typical example is the selection bias. The problem in these cases are that selection (determing who is treated or observed) and outcomes may be correlated. 2. Another is to identify the true coefficient in a linear regression when regressors are measureed with error.

2.2.1 Point Identification

We start by assuming some information ϕ is knowable. A simple definition of point identification is that a parameter θ is point identified if, given the model, is uniquely determined from ϕ. Notice that this definition of point identification is recursive in some sense. To identify θ, we first need to assume some ϕ is knowable, which means ϕ itself is identified. This identification of ϕ can only be justified by further assumptions of DGP (Data Generating Process).

For example, for a model $Y = X\theta + e$, we assume that $E(E^2) \neq 0$ and $E(eX) = 0$, and suppose ϕ includes the second order of (Y, X). Then we can conclude that ϕ is point identified, given by $E(XY)/E(X^2)$. Notice that the identification comes from the *assumptions* of model.

One common DGP is IID. Under this DGP, we can consistently indentify the distribution of observation W. Another DGP is where each data point consists of a value of X chosen from its support, then we randomly draw Y conditional on X, which is independent from other draws conditional on this X. Under this DGP, we can consistently identify $F(Y|X)$. We can also use more complicated DGPs, for example, we generally assume only the second order moments are knowable in time series. One reason is this being sufficient for identification, another reason is higher order moments become unstable over time. Assumptions over GDP are always needed, even in experienment data, and which specific assumptions to take depend on the model.

2.3 Low and Meghir, 2017, JEP

Low, H. and C. Meghir (2017, May). The Use of Structural Models in Econometrics. Journal of Economic Perspectives 31(2), 33–58

Defining a structural model:

A **fully specified model** make explicit assumptions about the economic actors' objectives and their economic enviroment and information set, as well as specifying which choices are being made within the model. They allow a complete solution to the individual's optimization problem as a function of current information set. Fully specified models are particularly useful in understanding mechanism of a policy, especially when we want to estimate some long-term effects of the policy.

A **partially specified model** relies on a sufficient statistic that summarizes choices not being modeled specifically. For example, assuming that the choices is only intratemporal instead of intertemporal.

Treatment effect models focus on identifying a specific causal effect of a policy while saying least about the theoretical environment. The pro is the cleaness of causality. The con is the limitation in exploiting the results outside. The identification of treatment model depends on assumptions that the experienment has not been compromised and there is no spillovers from the treatment units.

A combination of fully speficied model and randomized experiments can enhance anaysis for both. Experimental evidence can be used either to validate a structural model, or to aid in the estimation process (in identification).

Solving structural models

This has been described in Adda and Cooper(2003) well. The general process is: 1) write down the bellmand function; 2) discrete the state space and decision space; 3) use value function iteration to solve the bellman function.

Chapter 3

Moment Inequality

3.1 PPHI, 15 EMCA

Chapter 4

Others

4.1 Gentzkow-Kelley-Taddy, 17 WP

Gentzkow, M., M. Taddy, and B. T. Kelly (2017, February). Text as Data. pp. 1–53

An survey paper on how text analysis can be used in economic research. The main difficulty is the high-dimensionality of text.

Steps to reduce dimension:

1. Represent raw text $\mathscr{D}$ as a numerical array C
   (a) Divide $\mathscr{D}$ into individual documents $\mathscr{D}_i$, the level is determined by the level of V. For example, daily attributes or monthly attributes.
   (b) Feature Selection. Drop out punctuations, numbers, HTML tags, proper names and so on. Both very common and very rare words will be excluded. And words with the same stems will be reduced to one.
   (c) Limit dependence between words using N-gram. The idea is to only consider consider phrase of length n.
2. Map C to predicted values $\hat{V}$ of unknown outcomes V
3. Use $\hat{V}$ in subsequent descriptive or causla analysis

Reading not finished. Grimmer and Stewart (2013) also talks this problem. worth reading

4.2 Gillen-Montero-Moon-Shum, 18 MP

Gillen, B. J., S. Montero, H. R. Moon, and M. Shum (2018). BLP-Lasso for Aggregate Discrete Choice Models of Elections with Rich Demographic Covariates

This paper uses LASSO to pick instruments for BLP style model. They apply it in Mexico voting model.

Applications:

- Maybe in education choice

4.3 Seminar Notes

AI and Economics (`http://taddylab.com/slides/EconAI.pdf`)

This slides procides some clear intro on why and how AI (also deep learning) can be applied in economics. Many references provided in slides as well.

Not understand much. Needs further reading.

Is there any cases where only structural estimation can work, while reduced estimation, no matter how frequently A/B test, can not work?

Lixiong Li in IO brownbag introduces a way for identification where there is moment restrictions. The idea is to replace the specific distribution assumptions of the error term by some moment restrictions. By this, we can not have point identification in many cases, but can get set identification even with very weak conditions.

I think it a good paper.

Part II

IO

Chapter 5

Data and Techniques

5.1 Data Source

Uber and Taxi

- New York Taxi and Limousine Commission Trip Data
  - http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml
  - Literature: Bian Bo (JMP)

Nielson Consumer Panel Dataset

- Penn State has institutional access

ISMS Durable Goods Dataset 1

- Ni, J., S. A. Neslin, and B. Sun (2012, November). Database Submission—The ISMS Durable Goods Data Sets. Marketing Science 31(6), 1008–1013

Bureau of Transportation Statistics

- https://www.transtats.bts.gov

Airbnb data

- Penn State Department of Real Estate (?) how to get it

5.2 Timber Data by Forest Service

Timber Industry:

- Dropping off timber, then process timber to log, then process log to lumber by mill
  - Loggers do not possess mill machines, so they will sell the dropping timber to mills. At the same time since there are many different kinds of logs (e.g. in diameters), one specific mill may lack the machine for some logs and will resell them in the market. So the resale market after timber auction is frequent. This create a pro-collusion environment.
- Timber tracts are very different from one to another. For older timber tracts, the values of bidders are more heterogeneous. For young or second-grow timber tracts, the value is easier to estimate and we can think it more like common value auction.

Databook:

- BDA: number of bidders bid excess reserve price. In sealed auction, must bid larger than reserve price. In ascending auction, must actively bid at least once in auction.

  BDT: number of bidders actually show up in an auction.
  - e.g. BDA 0, BDT 1: 1 show up, bid lower than reserve price. Notice that BDA=1 does not mean the auction is not successful.
    - See 884F060103002, where BDA=0, BDT=1, the bidding value is the same as advalue, but the status (stt) is 1.
  - e.g. BDA 3, BDT 6: the price might increase too fast, the remaining 3 don't have a chance to bid.

Research Questions:

- Athey et al. (2011) shows that small bidders (loggers) pay higher in sealed auction. Why they do not move to the ascending auction in this case? Why the arbitrage does not work here?
  - Maybe it's easier to collude in ascending auction than in sealed auction. Then independent bidders will prefer to stay at sealed bid auction even the realized price is higher. Because if they move to ascending auction, then they are likely not to win.
- What if the reserve price is strategically picked by the auctioneer?

- Timber auction is multidimensional. Each tract have different species, and each firm will have different values over this species. Can we use the Nima's optimal multidimensional mechanism design to verify the data?

- Timber firms may have special interest in some species, so they may want to participate in specific auctions. And thus the private value assumption does not hold.

  - Bob: Firms have strong incentive to collude in this case. Firm A prefer species a, firm B prefers species b. Such preference may derive from their specialization in processing, and is commonly known among bidders. Then in an auction where there are 80% specie a and 20% specie b, it is very likely that firm A will win and sell species b to firm B after auction.
  - Then they are very likely to be engaged in an collusion. How can we distinguish it from the pure collusion based rotation mechanism, where firm A will attend in some auctions more frequently while B will attend others?

5.3 FCC data

Related Literature:

- McMillan (1994)

- Cramton and Schwartz (2000) clarifies many terms in a narrative way.

- Weber (1997) introduces *strategic demand reduction* in simultaneous ascending bidding auctions.

  - e.g. Two firms, three objects. Each firm either any one object 100, and any combination 200. Then if one firm take the strategy that keep bidding on two objects until some of which exceeds 100. Then the other firm will find it optimal to bid at the other object. The auction will end at a very low bid for all three objects.

- Krishna and Rosenthal (1996) discusses payoff complementarity in ascending bidding auction. Their model is based on second price sealed auction.

5.4 Techniques

High Resolution Satelite Data:

- Pixel is as accurate as 0.5m, much finer than the Night Light data;

- Use machine learning to identify the roads, cars, buildings etc;
- Jonathan Hersh has a paper over this. **Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being and Geographic Targeting**

Google Street View Data:

- Gebru et al. (2017) uses Google Street View data to collect automobile information, and use machine learning to predict socioeconomic conditions of different cities.

Paralelle Computation:

- A detailed cook manual on Paralelle computation of RBC model:

  `https://www.kellogg.northwestern.edu/researchcomputing/docs/ParallelComputing/ParallelizationForRBC`
  `pdf`

- Another source: `https://bfi.uchicago.edu/sites/default/files/file_uploads/Lecture_Parallel_Programming_Chicago%20%282%29.pdf`

Text as Data:

- Course Materials by Justin Grimmer
- Grimmer, J. and B. M. Stewart (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political Analysis 21(3), 267–297
- Gentzkow, M., M. Taddy, and B. T. Kelly (2017, February). Text as Data. pp. 1–53

Chapter 6

Cartels and Collusion

A very good narrative of cartel and bidding rings is from *Economics of Collusion* by Marshall and Marx. I will briefly summarize.

- A cartel in most cases will need explicit agreement instead of just tacit collusion.
- They will hire a consulting firm to negotiate the meeting and to keep track of all records.
- They will fix a market share instead of just prices, and use 'true-up' to redistribute as agreed at the end of the period.
- They will also change the incentive of the sales team to avoid over selling.

A bidding ring is a group of bidders where they will not compete within the group. There are two benefits. First, competition is directly lowered and price would be lower for bidders. Second, by reduce bidding within the group they will not leak information to other less informed bidders, and reduce competition indirectly.

Measure of Concentrations:

- Herfindal Index:

6.1 Green and Porter, 84 EMCA

Green, E. J. and R. H. Porter (1984). Noncooperative Collusion under Imperfect Price Information. Econometrica 52(1), 87–100

Green and Porter (1984) points out that **price war can be a part of collusion equilibrium**, challenging the view that price war being a signal of competition. The key point is that there is **uncertainty in demand**, which is after the decision of firm's quantity decision. So firms are unable to distinguish between a low demand shock and someone's possible deviating.

They show that the equilibrium strategy is a **trigger strategy with forgiveness**. For realized price p higher than some threshold $\bar{p}$, firms choose the cooperative quantity; otherwise they will choose the Cournot quantity for T periods, where the punish time is given (maybe by explicit negotiation process).

The dynamic incentive is very interesting here. In equilibrium, no firms will have incentive to deviate. So the only possibility for a low price is a bad demand shock. Despite everyone knowing this, they will still choose the Cournot quantity for T periods as a punishment. This is to give correct dynamic incentive. Suppose a strategy profile features no punishment even in low price. Then firms would have incentive to output more since they would not be punished anyway.

Learning Process:

One thing I feel uncomfortable is the information structure of the firm. In this paper, firms are assumed to know the realization of price, the distribution of shocks θ, and the function form of price $p()$.

If they do not know the distribution, not do they know the functional form. Then it would be impossible for them to distinguish the volatility in θ and in quantity. Thus it would be impossible to detect probable deviation. [a]

[a] In the paper, we can detect deviation at some probability. But even this cannot be satisfied because the distribution is unknown and relative parameters are unidentified.

Identification: in appendix they show that in principle the collusion behavior can be identified. But in practice it is generally quite difficult.

6.2 Rotemberg-Saloner, 86 AER

Rotemberg, J. J. and G. Saloner (1986). A Supergame-Theoretic Model of Price Wars during Booms. American Economic Review 76(3), 390–407

Rotemberg and Saloner (1986) also tries to explain price war, but they show that **price wars are more likely to happen during the demand boom instead of doom.** In their model, firms make decisions *after* realization of shocks. So firms can distinguish between the price drop by deviation and price drop by low demand. Then their result is quite intuitive. It is more profitable to deviate when demand is high. So the collusive price at high demand should be lower, the lower is collusive price, the less attractive to deviate.

Comments:

- With more specific assumption on knowledge of firm, they can say more about the competition structure. In Green and Porter (1984), the only reasonable competition mode is Cournot. But both Bertrand and Cournot are appropriate in this paper.
- The establishment of price competition equilibrium is skillful. They first assume a exogenous punishment to generate optimal decision, then use such decision to generate an actual punishment. This constitutes a fixed point argument. [1]
- This model is '*really about countercyclical pricing—firms have perfect information and adjust prices smoothly in response to demand conditions.*'(by Ellison (1994) p.38). There is no *regime shift* like in Green and Porter (1984) since the past has no uncertainty.
  - It is not important whether the future has uncertainty. Say, the demand is seasonal and is certain. It won't change the result, the pricing will still be countercyclical. Intuitively, what the firms compare is the current deviation profit and expected future gain, thus future uncertainty is not important (as long as firms are risk neutral).

6.3 Porter, 83 Bell

Porter, R. H. (1983). A Study of Cartel Stability: The Joint Executive Committee, 1880-1886. The Bell Journal of Economics 14(2), 301–314

Porter (1983) is specifically to test Green and Porter (1984) using JEC data. [2]

Estimation Strategy:

- **Goal** is to test whether regime switch happens in supply side, and to identify when such switches happen. This paper does not test whether firms actually play the equilibrium as described by Green and Porter (1984), but only describes the reversion pattern.
- **Data** is aggregate time series price and quantity.
  - KZ: notice there is some internal inconsistency in this setting. Porter assumes 1) no secret deviation, without giving justification; 2) homogeneous good, which is largely due to data limitation; and 3) firms do price competition. But then we should see prices near marginal cost, which should not fluctuate a lot.

[1] Same trick as in Peters and Siow (2002) small market case.
[2] Joint Executive Committee, which is widely used since it records the cartel in 1880-1886 for the train industry in US.

- **Firm Strategy:** The firms in JEC actually use price instead of quantity as strategic variable, and they use differentiated products (through different service). This is different from Green and Porter (1984) model.
  - But in econometrics model part, this paper seems to still assume the quantity competition. This may due to the data limitation — they only have aggregated price and quantity.
  - Firm's decision problem is summarized by

$$p_t(1+\theta_{it}/\alpha_t) = MC_i(q_{it})$$

  Porter notes that this form can encompass Bertrand competition, Cournot competition and perfect collusion, with θ_{it} being 0, $(0,1)$, 1 respectively. This seems reasonable, what I don't understand is how it can be compatible with Green-Porter's model? In their model it is obvious that Cournot is the only possible way of competition.

  This form can be derived from firm profit maximization. Suppose

$$\pi_i(q_i, q_{-i}) = p(q_1, \ldots, q_n)q_i - C(q_i)$$

  by taking derivative w.r.t. q_i we will get:

$$p(1+\frac{1}{\varepsilon_{q_i,p}}) = MC(q_i)$$

  Also notice that this can actually encompass both Cournot and Bertrand. In Cournot, we have $p(q_1, \ldots, q_n) = p(\sum_{i=1}^n q_i)$; and in Bertrand we have $p(q_1, \ldots, q_n) = p(\sum q' | q' \in \max\{q_1, \ldots, q_n\}, 0 = q_i \notin q')$

  - In firm optimization problem, there is no uncertainty. But in turning it to reduced form, the paper has to add error term U_{2t} to match the data. In adding this term, we implicitly assume 1) firms do not know this shock when making decision; 2) this shock is aggregate for all firms (?).
- **Estimation:**
  - Demand Curve:
  - Supply Curve:
  - Simultaneous Equation System:
  - Conditional pdf:
  - Unconditional pdf:
  - Convergence algorithm:

6.4 Genesove-Mullin, 01 AER

Genesove, D. and W. P. Mullin (2001, June). Rules, Communication, and Collusion: Narrative Evidence from the Sugar Institute Case. American Economic Review 91(3), 379–398

Genesove and Mullin (2001) narratively describes how the sugar cartel in 1930-1936 actually works. They find cartel's action deviates from what collusion theory predicted. The main idea they try to convey is: **collusion is more a contraction problem than a market structure problem**.

Questions they want to answer:

- How firms coordinate their pricing?
  - Mainly through rules, and by ordinary meetings to adjust to rules. Such meetings are necessary due to the uncertain nature of competition environment.
- How prevalent is cheating? What's the punishment?
  - Cheating is not discussed in neither Green and Porter (1984) nor Rotemberg and Saloner (1986). There the price war is just an equilibrium result, and no one has incentive to 'secretly cheat'.
  - But in reality cheat actually happens. Because contract is always incomplete and firms have many tricks to escape. But retaliation does not always happen. They will first make some effort to partially identify a cheat and a bad draw, e.g. by asking the suspect to show some hard evidence. When the cartel is convinced that someone has cheated, they are mostly likely to take tit-for-tat strategy instead of trigger.

Below are notes taken during the reading.

- Cartel mainly collude through rules.
  - Open prices and publicly announced items
  - Prior notification
    - Complementary to open announcement. Ensure that firms can potentially retaliate before consumers react.
  - Restrictions on contractual practices between the refiners and downstream firms
    - Mainly to eliminate discriminatory pricing
    - Very detailed and standardized. Because 'every contractual term can mask a price cut'. The standardized model is implicitly assumed in Green-Porter, though homogeneous good and the extremely simplified contract and strategy form.

    * Such rigidity is also costly to cartel
  - Alternative explanation for such limitation on contracts:
    * Limiting non-price competition
    * Authors did not agree. They argue that the *contractual harmonization* conveys more than non-price competition limitation.

- Collusion contract is always incomplete
  - The regular and frequent meetings are to fix the holes
- Why firms do not directly retaliate when they find cheating?
  - Different from Green and Porter (1984), firms in Sugar Cartel can argue within the institute meeting. Such meeting gives them a chance to figure out whether it is a shock or a intentional cheat. In other words, firms do have a choice to reduce the confounding uncertainty, which reduces their incentive to retaliate.
- Firms do not retaliate to every cheating. They do not use massive retaliation unless the cheating is very high. The retaliation they take generally is 'matching', kind of tit-for-tat.
  - One reason to desist from full-scale retaliation (as in Green and Porter (1984)) is the vertical contractual arrangement in the industry. Firms sell sugar through brokers. Brokers will have much higher incentive to deviate since their number is much larger than that of firms. Such structure made deviation more prevalent. It is also more costly, since one more layer of brokers make it more difficult for firms to switch between two pricing regimes.
- Firms rarely react to market share.
  - Market share is too noisy a signal.
- Firms choose to meet frequently to adapt quickly to external change, and to fix any holes in rule as soon as possible. They seem not care much about the *renegotiation problem*.

6.5 Graham-Marshall, 87 JPE

Graham, D. A. and R. C. Marshall (1987). Collusive Bidder Behavior at Single-Object Second-Price and English Auctions. Journal of Political Economy 95(6), 1217–1239

This paper models the collusion within bidders in Second Price Sealed auction and English auction in IPV framework. They find collusion strategy equilibrium can be stable in such settings. Coalition of any size is stable, and the expected payoff to each member increases with the coalition size.

- In second price auction. The optimal reserve price is a increasing function of coalition size.
- In English auction. The optimal reserve price is identical to that of second price auction, and independent of observed highest bid. This result is striking since it seems at first that auctioneer at English auction can do better with this additional information.

Some stylized facts are not incorporated in this model:

- Fact: brokers will not be invited to the coalition. If a broker has a customer, his willingness to pay is much stronger than any dealer. Thus he does not want to join the ring since other ring members will push price high in the knock-out round. If a broker does not have a customer, the he is not really coming to bid but to share the 'ring hat'. The ring does not want him in this case.

  → This fact cannot be encompassed in this model. Because the model assumes IPV, while brokers represent a very special distribution (0 or very high) of value, against the assumption.

6.6 MMRS, 94 GEB

Marshall, R. C., M. J. Meurer, J. F. Richard, and W. Stromquist (1994). Numerical Analysis of Asymmetric First Price Auctions. Games and Economic Behavior 7(2), 193–220

This paper raises a algorithm to numerically solve equilibrium in heterogeneous auctions. And they use this algorithm to investigate simulate some facts for auction coalition in first price auction.

Facts from Simulation:

- Revenue equivalence of first price auction (or Dutch) and second price auction (or English) does not hold in existence of collusion. The expected revenue is higher in first price auction.
- In second price auction, the individual has no incentive to deviate from the coalition even when coalition size is very large. Actually the surplus of coalition gain to individual gain increases with coalition size k_1.
- In first price auction, the individual has no incentive to deviate when coalition size is small, but does have incentive to deviate when coalition size is large. The relative advantage of individual bidder increases with coalition size k_1.
- In ascending auction, for any given value, the weaker bidder will be more aggressively than stronger bidder.

  - This induces inefficiency. Stronger bidder with higher value may lose the auction.
  - Payoff equivalence not holding in this environment?
    * maybe sealed auction and ascending auction does not reduce to the same direct mechanism in this environment

- The same features are also found in Athey, Levin, and Seira (2011). Although in Athey el al., the heterogeneity is not from collusion but from different cost distribution between large (mill) and small (logger) bidders. But the mathematical representation is the same, larger bidders value distribution FOSD that of the small bidders.
  - They find the timber auction has following stylized facts: sealed first price auction attract more small bidders, shit the allocation toward these bidders, and can generate higher revenue to auctioneer.

Chapter 7

Price Dispersion

7.1 Butters, 77 RES

Butters, G. R. (1977). Equilibrium Distributions of Sales and Advertising Prices. The Review of Economic Studies 44(3), 465–491

This paper tries to generate price dispersion in a homogeneous environment. Homogeneity is in the sense that commodity characteristics, consumer preference, seller cost are all the same. The only heterogeneity comes from seller's strategy on pricing and advertisement.

Unlike previous literature only focusing on consumer side, Butters models both consumer search behavior and the firm strategy. In other words, previous models are 'partial-partial equilibrium', while this paper is 'partial equilibrium'.

No Consumer Search:

- Seller: decide the amount of ads, for each ad can post a different price. Each ads costs b.
- Buyer: the only way to purchase is through ads, and pick the lowest price (unless all prices higher than reserved price m). Each ads reach each consumer with equal probability, that's to say, the probability distribution of number of ads of any given consumer is Poison distribution. Most importantly, the probability that a consumer receives no price quotes lower than p, given the ads distribution $A(p)$, is $exp(-A(p))$.
- The model is solved purely at market level. Denote $A(p)$ to be the number of ads with prices equal or

lower than p divided by consumer number M. $S(p)$ to be the sales actually happen with price lower than p divided by consumer number M. $\pi(p)$ be the probability of selling a good by pricing at p.

- We thus have $s(p) = a(p)\pi(p)$, where $s(p), a(p)$ are derivatives of S, A respectively.
- The profit function is:

$$\text{profit} = \int [(p - p_0)\pi(p) - b]a(p)dp$$

- By no deviation condition, the marginal profit at any equilibrium price level should be zero. Thus, $(p - p_0)\pi(p) = b$, and $\pi(p) = b/(p - p_0)$
- By similar arguments as in allpay auction, the equilibrium price can only be a range $[p_0 + b, m]$
- IMPORTANT: $\pi(p) = exp(-A(p))$
  - $\pi(p)$ is the probability that an ad with price p being accepted. This is the probability that this ad goes to a consumer with ZERO price quotes lower than p. By our previous analysis, $A(p)$ is the number of ads per consumer, i.e. the single time probability that a consumer receives an ad with price lower than p. Then by the results of Poison distribution, the probability that a consumer receiving no price quotes lower than p is $exp(-A(p))$.
  - This is important since it connects the individual behavior and market index. Firms only cares about π when making ads decision; while $A(p)$ can induce $\pi(p)$. These two have to be consistent in equilibrium.
- The equilibrium is easy to solve with above characterization.

With Consumer Search:

- The setting is quite tricky. The equilibrium is greatly simplified due to the following two assumptions of *word-of-mouth search*:
  - ASSP-1: The probability that a search yield an offer from any particular seller is proportional to his share in total shares;
  - ASSP-2: The probability that this offer is at a price within a particular interval (p_1, p_2) is proportional to the seller's sale at this particular interval.

  Above two assumptions makes the profit from searching consumers exactly proportional to that of non-searching consumers. So we only need to maximize the profit from non-searching consumers, which has the same structure as no-search case.

  - KZ: I find these two assumptions uncomfortable. This is not about the search behavior, but all about the optimal advertisement.

- Claim: consumers will search iff they receive no ads, and in this case they will accept the first price they find.
  - The second half is direct implication of the assumptions, especially assumption 2.
  - For first half, denote the highest posted price p_{max}, and the threshold price of searching p_c. Obviously $p_{max} \leq p_c$. Further we claim that $p_{max} = p_c$, otherwise the firm can increasing highest posted price by ε while keep advertisement same as in p_{max}.
- The equilibrium is very similar to no-search case. The only difference is $\pi(p_{min}) = 1/\phi > 1$. At lowest posted price, one ad will induce more than one purchase.
  - This is because now we have $1 - \phi$ portion of people receive no ads, at any price interval. Thus conditional on the consumer receiving at least one ad, p_{min} will for sure lead them to buy. But such ads will induce even more purchase because those receiving no ad will also buy from this seller at this price p_{min} proportionally.
  - This makes sense because since consumer can search, firms can transfer the load of advertisement partly to the consumer side.
- The portion ϕ can be pinned down in equilibrium by the fact that eventually every consumer will buy.

7.2 Sorensen, 00 JPE

Sorensen, A. T. (2000). Equilibrium Price Dispersion in Retail Markets for Prescription Drugs. <u>Journal of Political Economy</u> <u>108</u>(4), 833–850

Facts and Identification Strategy:

- This paper is a reduced form paper testing price distribution in drugs.
- The identification strategy is to **use *frequency* as the key explanation variable**.
  - This is the *comparative static* approach. Because by Butters (1977) the price dispersion will shrink if the cost of search is lower. Here the searching frequency is a proxy as search cost.
- Measure of price dispersion is **price range**. This is an absolute measure, because search cost generally do not vary with price scale. There are other measures possible, for example standard deviation, or relative range.
- The environment is carefully picked for clean identification:

  - The market is some small isolated markets, with sufficiently many pharmacies within each market.
  - The frequency of purchase is determined by instruction of doctor, which is uncontrollable by drug consumers.
  - The drug price is posted, thus is in principal public so that econometricians can collect them. But consumers have to pay the search cost to gain such information.
  - All pharmacies face similar cost.
- They also exclude the possibility of heterogeneous pharmacy service.
  - First, they show that pharmacies have very different rankings for different drug price. This fact is inconsistent with the explanation of pharmacy level heterogeneity.
  - Second, they run a residual regression to show that pharmacy heterogeneity at most account for 30% of price dispersion.

What Cannot be Answered:

- The data generating process. How exactly search cost affects the price dispersion.
- Quantifying search cost.
- This has to be done with a structural model like Hortacsu and Syverson (2004).

7.3 Hortacsu-Syverson, 04 QJE

Hortacsu, A. and C. Syverson (2004, May). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. The Quarterly Journal of Economics 119(2), 403–456

Facts and Explanations:

- Mutual fund prices are hugely dispersed, despite their financial homogeneity. The 75th to 25th price ratio of Index Funds is 3.1;
- Nonportaforlio Differentiation. Funds attributes like manager tenure, the age of the fund, rate of taxable distributions will not directly affect return, but do affect investors' choice.
  - Only focus on vertical differentiation and assume homogeneous preference over these attributes. Because horizontal differentiation is unidentifiable from search cost differences.

- Search Cost.
- Switch Cost. If switch cost is important, then we should observe in data that inflow/outflow of S&P Fund in a non-load funds family should be less volatile than those in a load funds family. Because when index fund performs poor, those in a load fund family will switch to another fund in the same fund family, inducing larger imbalance within the fund family. This is not significant in data.

Model:

- Funds are different in both prices and other vertical attributes. But investors have homogeneous utility form. [1] We assume a specific utility form, where price enters linearly:

$$u_j = W_j\beta - p_j + \varepsilon_j$$

- Investors have cost c_i for each search, which is distributed in population as $G(c)$. An investor searches sequentially until the expected utility gain is smaller than search cost:

$$c_i \leq \int_{u^*}^{\bar{u}} (u - u^*)dH(u)$$

  where $H(u)$ is investors' belief about the distribution about funds' indirect utility, which is unknown to econometricians and have to be assumed. We assume that funds provide N levels of utility, $u_1 < ... < u_N$. If we further assume equal sampling probabilities, we have:

$$H(u) = \frac{1}{N}\sum_{j=1}^{N} I[u_j \leq u]$$

  Investors are assumed to know the value of each u_j, they are only uncertain about which fund provides which utility level.

  - KZ: this assumption very strong, which greatly reduces the dimension of information investors need to learn about.

- The optimal search rule yields critical points in search cost:

$$c_j = \sum_{k=j}^{N} \rho_k(u_k - u_j)$$

  where ρ_k is the probability that fund k is sampled on each search, which is known to all investors. The RHS is the expected payoff of another search conditional on already found j. Thus c_j is the highest

[1] We cannot identify taste heterogeneity using BLP. The taste differentiation is unidentifiable with search cost differentiation.

possible search cost that could allow such search.

- KZ: ρ_k is just the density of $H(u)$. If equal sampling, then $\rho_k = 1/N$ for all k. If not equal sampling, it should be conditional on how much time an investor has already searched, unless we assume the funds pool is infinite.

Identification:

- **Goal:** identify the search cost and coefficients of other attributes with only price and market share data.
- Funds market share can be written in cost thresholds. Thinking of u_1 with lowest indirect utility. Only those with really high search cost will end up buying fund-1, and they have to be unlucky to draw this fund at the first time.

$$q_1 = \rho_1(1 - G(c_1)) = \rho_1\Big[1 - G\Big(\sum_{k=1}^{N} \rho_k(u_k - u_1)\Big)\Big]$$

For fund-2, there are two sources. One investor may have cost higher than c_1 and draw fund-2 in the first time; or he has cost $c_2 < c < c_1$ and draw fund-2 first time or directly after fund-1. [2]

$$q_2 = \rho_2(1 - G(c_2)) + \frac{\rho_2}{1-\rho_1}(G(c_2) - G(c_1)) \tag{7}$$

Similarly, we can write N equations for all funds market share.

  - Thus we can back out $G(c_j)$ with known ρ_j and q_j for all j. Moreover, we can normalize $G(c_N) = 0$ because we have $c_N = 0$ by threshold equation, and search cost have to be nonnegative.
  - We can take derivative with p_j, because c_j is influenced by u_j which contains p_j. Thus we can write $N - 1$ derivatives:
$$\frac{\partial q_j}{\partial p_j} = f(\rho_1, ..., \rho_N, g(c_1), ..., g(c_j)) \tag{10}$$
where $g(c_j)$ is the p.d.f. of cost distribution at c_j.

- For supply side, we first write the FOC of fund j.

$$q_j(p, W) + (p_j - mc_j)\frac{\partial q_j(p, W)}{\partial p_j} = 0 \tag{9}$$

In above equation, only q_j and p_j is observable to econometricians. If we further assume Bertrand-Nash competition, we have:
$$\frac{\partial q_j(p)}{\partial p_j} = -\frac{q_j(p)}{p_j - mc_j}$$

[2] $Pr = \rho_2 + \rho_1\rho_2 + \rho_1\rho_1\rho_2 + ... = \rho_2/(1 - \rho_1)$

If we know mc_j, then we can calculate $\partial q_j/\partial p_j$, and plugging into (10) get us $g(c_j)$ for all $c_1, ..., c_{N-1}$

The paper says mc_j can be estimated, don't know how

- Knowing G_j and g_j for all $c_1, ...c_{N-1}$, we can back out all c_j through:

$$G(c_{j-1}) - G(c_j) \approx 0.5[g(c_{j-1}) + g(c_j)](c_{j-1} - c_j)$$

Recall that $c_N = 0$, we have known all c_j. Since we know all c_j and $G(c_j)$, we have backed out the distribution of search cost (in these thresholds).

Estimation:

- **Basic Model:** homogeneous funds except for price, equal sampling $\rho_j = 1/N$
  - By homogeneity, we have $u_j = u - p_j$. Thus c_j is only function of p_j's.
- **Homogeneous Fund + Unequal Sampling Probability:**
  - By homogeneity, we have $u_j = u - p_j$. Thus c_j is only function of p_j's. Since we know p_j for all j, we then know c_j as well.
  - By unequal sampling, assume specifically:

$$\rho_j = \frac{Z_j^\alpha}{\sum_{k=1}^N Z_k^\alpha}$$

    where Z_j is the attributes that will affect the likelihood of being found, e.g. advertisement expenditure. We use age of a fund as proxy.
  - Estimate the model by fitting $(N-1)$ (7) and $(N-1)$ (9). We know c_j, Z_j, and $G(c_j)$ can be written as function of Z_j and unknown parameter α. So we have $2N-2$ equations to estimate α and mc_j for all j.
- **Heterogeneous Fund + Equal Sampling Probability:**
  - Since funds are heterogeneous, c_j cannot be directly derived from u_j. But we can still non-parametrically identify c_j as described in previous section.
  - For identification, we must assume equal sampling probability. We also assume mc_j is constant and equals 10 basis point for all j.
  - Estimation result shows that overall search cost decreases sizable amount in 1996 to 2000, but in upper quantile increases. This may be due to the entering of many novice family investors.

Questions Remaining:

- The assumptions of drawing funds with replacement is not satisfying, especially when the industry is very concentrated in several star funds.
- If we want to investigate both horizontal differentiation and vertical differentiation. Then we have to know not only the individual level data on what choice they made, but also the *choice set* they have.
  - There are two possible interpretation for a fund to be chosen. 1) the consumer observes all the funds and prefers this fund most; 2) the consumer only observes a limited number of funds and have to pick one. This causes the identification problem between preference and search cost.
- Bundling problem. The consumers do not just buy one fund, but buying a bundle of funds. They react to the whole bundle price instead of single fund price. This is one specific kind of product differentiation.
  - The importance of such product differentiation is, it can hardly say that such differentiation is only vertical. Different investors have different financial goals, and thus have different preference on bundles.

7.4 Allen-Clark-Houde, 14 AER

Allen, J., R. Clark, and J.-F. Houde (2014, October). The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry. American Economic Review 104(10), 3365–3396

Empirical Takeaway:

- The competition of firms mostly benefit consumers with low search cost.
  - The merger led to an increase in *average* interest rate of appx 6 bps.
  - This merger effect is between 7 to 9 bps for consumers with lower and middle range percentile search cost; and has no effect for consumers with top 30% serch cost.
  - The merger led to a 16% decrease in interquartile price range.

Identification and estimation part to be added.

Baye, M. R., B. De los Santos, and M. R. Wildenbeest (2013). The Evolution of Product Search. <u>The Journal of Law, Economics Policy</u> <u>9</u>(2), 1–21

- Price comparing sites becoming less popular, more consumers choose to directly go to like Amazon and eBay. These websites also have 'marketplace'.
- Many searching behavior are not observed. Non-texture data (e.g. in the menu, directories).
- Search time across platforms can hardly be compared. Because maybe it's enough to search only once in Amazon, but many times in eBay.

Random Ideas

- Collusion when there is consumer searching cost. Intuitively, collusion should be easier to sustein with [no idea]
- Awaya-Krishna (2016)

Chapter 8

Auction

8.1 McMillan, 94 JEP

McMillan, J. (1994). Selling Spectrum Rights. Journal of Economic Perspectives 8(3), 145–162

- Lessons from New Zealand and Australia:
  - Second price auction may have political defect and revenue loss. People are dissatisfied to see firms pay less than their revealed true value. And reserve price will enhance revenue when bidder number is small.
  - Australia's auction shows that penalty for default is necessary to restrain incentive of over bid.
    - FCC: the highest bidder need to pay the difference between his bid and actual sale price if he chooses to exit.
- Open Bidding v.s. Sealed Bidding
  - Open auction can mitigate winner's curse because bidders can learn from others' biddings, and thus be less cautious.
  - Sealed bid auction increases revenue when bidders are risk averse. Because bidders will bid higher to ensure winning in sealed auction.
  - Sealed bid auction deters collusion.
  - FCC: run multiple rounds of sealed auction, announce bids after each round, but do not announce bidder identity.

- Simultaneous v.s. Sequential Auction
  - Sequential bidding have two disadvantages:
    * It is less flexible, impeding the possible aggregation. The firm cannot go back and adjust his portfolio.
    * Encourage predatory bidding in early stage.
  - Simultaneous auction has difficulty in implementing effective stopping rule.
  - FCC: for 20% core licenses, use simultaneous stopping, and use license-by-license stopping for the rest.
- Combinational v.s. Non-Combinational Bidding
  - Combinational bidding allows bidders to bid for bundles of licenses together. It is efficient in the sense that different licenses are complementary and have economics of scale.
  - Combinational bidding has another advantage of expanding the preference space.
  - Combinational bidding will refrain competition. Because there is free rider effect within independent bidders. Even there total value of separately awarded licenses exceed that of the nationwide bidder, they would not like to bid high because they only get a share of winning rent.
  - FCC: combinational bidding not allowed in simultaneous bidding; allowed in sequential bidding with a possible premium.
- Aid Designated Bidders
  - Give those bidders a price preference.
- Royalties v.s. Single Payment
  - Royalty partially shifts the risk to government, thus increasing bidders' willingness to pay.
  - Makes auction more like common value. The paper says this will stimulate competition, don't understand.
  - Causes post-auction incentive problem.
  - Difficulty to isolate revenue induced by such license in practice.

8.2 Kawai, 18 WP

Kei Kawai (2018 WP): Missing Bids and Scoring Auctions

There are two main contributions for this paper:

- It introduces a way of measuring collusion level in auction.
- It empirically shows that scoring auction is more competitive (less collusive) than pure price-based auction.

Detecting Collusion:

- Define the distance of bid-i to the winning bid of remaining bidders to be:

$$\Delta_i = bid_i - \min_{j \neq i} bid_j$$

  If we pile (sum up) Δ_i for all bidders, and plot the density, we should find them to be relatively smooth. But in data, the authors find that the density around $\Delta = 0$ is a low spike. This suggests collusion.

- Intuition is, if there is no collusion, then the bidders can increase their bids by a little bit. Since the density of Δ near 0 is very low, the probability of losing the auction by increasing bid is low. So the optimal strategy should be to increase the bidding price.
- The author further introduces a scale measure of collusion possibility. Do not understand this part.

Score Auction:

- The auctioneer of score auction does not take price as the only deterministic attribute for allocation. He will also consider the *quality* of the bid. But since the quality is a multi-dimension property, there is uncertainty about the allocation result. This uncertainty is essentially uncertainty about auctioneer's preference.
- Such uncertainty makes it more difficulty to monitor deviation within a bidding ring, thus makes the cartel more difficult to sustain. The intuition is from Green and Porter (1984).

Chapter 9

Spatial Competition

9.1 DGT, 78 EMCA

d'Aspremont, C., J. J. Gabszewicz, and J. F. Thisse (1979). On Hotelling's "Stability in Competition". Econometrica 47(5), 1145–1150

This paper serves as an example of clear writing.

Hotelling's 'converging to middle' result is not a generic property of the model. For example,

- When transportation cost is linear, there exists no pure strategy Cournot Eqm when two firms are very close;
- When transportation cost is quadratic, there exists pure strategy Cournot for any location pair, but both firms would like to move away from each other.

Intuition is, for p_1^* to be equilibrium price against p_2^*, it has to be better than any other price. Specifically, it has to be better than a price $\hat{p}_1$ which will grab all consumers from firm 2. When two firms are located close, $\hat{p}_1$ will be attractive. However, this cannot be an equilibrium since firm 2 will do the same reasoning. This is like a double auction with a cap, here the cap is decided by the distance between two firms.

9.2 Chaterjee, 18 JMP

Chatterjee, S. (2018). Market Power and Spatial Competition in Rural India

This job market paper estimates the market power within spatial competition framework.

Settings:

- Market is divided into several regions. Within each region, there are many farmers and intermediaries. Farmers first sell good to intermediaries, then intermediaries sell them in the retail market.
- Farmers in one region can only sell to intermediaries in the same market. Intermediaries has no such restriction.
- All transactions are subject to iceberg cost. So each intermediaries enjoys some monopoly power, which is increasing with scarcity of intermediaries in one region.
- Farmers do Nash bargaining with intermediaries. The whole market is playing a Nash-in-Nash game. Define the optimal alternative by:

  $$\underline{p}(m) = \max_{k \in \mathscr{M} - \{m\}} \{\frac{p^f(k)}{\tau_{mk}}\}$$

  where τ_{mk} is the distance cost from m to k. And the Nash bargaining outcome is the solution to:

  $$\max_{\lambda}(\lambda - \underline{p}(m)q_m)^{\delta}(p^r_m q_m - \lambda)^{1-\delta}$$

- Void the region restriction will increase the welfare of farmers. Intuitively, if information gathering is of no cost, then enlarging farmers' information set will do them no harm.

KZ Questions:

- What if farmers have to search for the cost? Will the welfare result hold?
- This paper introduces a 'ripple effect' of competition spread through the neighbors. How we can modify this to our liquor data?

Chapter 10

Vertical Relationship

10.1 Horn and Wolinsky, 88 RAND

Horn, H. and A. Wolinsky (1988). Bilateral Monopolies and Incentives for Merger. The RAND Journal of Economics 19(3), 408

Highlights of the Paper:

- When price is determined by negotiation, the incentives to merger will be very different with the case where price is unilaterally set.
- Symmetric case. Two suppliers, two retailers, independent or simultaneous bargaining. When downstream firms are strategically complementary. Then a monopoly upstream firm is *less* profitable than two independent suppliers.
  - [Intuition] Because when a monopoly supplier has to take into account the profit of both downstream firms, which weakens its bargaining power when the downstream firms are strategically complementary.
- Asymmetric case. One supplier, two retailer, sequential bargaining. Whether the upstream firm will earn more profit through merging depends on the reference point (disagreement point). When the disagreement point is high, then the supplier can earn higher profit though sequential bargaining; otherwise profit is higher though simultaneous bargaining.

Model:

-

Pros:

-

Cons:

-

Chapter 11

Demand Estimation

Reference: Shum (2016) CH 1 is very good for motivation and techniques.

Motivation for demand estimation

- Firstly, it is important per se, we want to understand consumer behaviour. And we want to *predict the results of price change*, and *introduction of new products*. Such counter factual predictions can only be built upon a structural model.
- Secondly, estimation of demand function helps us better understand market structure. Markup itself is rarely observed (with some exceptions, like the electronic industry), but we can generally observe demand elasticity. Recall that $\frac{p-MC}{p} = 1/\varepsilon$ (see **??**). Estimation of ε will help us know the markup of firm, thus infer market structure.

Challenges in demand function estimation:

- Too many parameters. Think of a market with 100 commodities and we want to estimate the cross elasticities. The number of parameters will be 10,000. Obviously it is impossible to have enough data to estimate so many parameters.
  - Solved by nested logit model.
- Unobserved quality. This will causes the endogeneity of price. And because the unobservable quality affects the selling in a highly nonlinear way, so IV is not very applicable in this case.
  - Solved by BLP style estimation.

11.1 Brief Literature Review

There are two branches of demand function estimation. The early literature features differentiated goods (continuous choices). Classical examples include Hausman et al. (1994) introducing a nested approach to reduce the dimension of parameters.

Recent literature focus much more on discrete choice model. The seminal papers are Berry (1994) and Berry, Levinsohn, and Pakes (1995). The address of these two papers is to overcome the **endogenous price problem due to unobserved quality**. Suppose we use a characteristic approach, where consumer utility and thus their purchase decision is decided by characteristics of products. Some of the characteristics are obervable to consumer (and firms) but not to econometricians, for example the advertisement. For higher unobserved quality, the equilibrium price will be higher, leading to a downward bias of price elasticity.

Berry (1994) comes with a method to overcome this endogeneity problem with only *market level data*.

- The basic idea is a IV based two step estimation.
- First step: denote δ_j to be the 'mean utility' for brand j. Notice this includes unobserved quality. Under some (weak) conditions, δ_j can be *inversed* by market shares $(s_1, ..., s_J)$, and thus we can get $\hat{\delta}_j$ by observed market shares. This inversion step is the key for this paper, and many following papers treating endogenous quality problem.
- Second step: By definition, $\delta_j = X_j\beta - \alpha p + \xi_j$. Thus we can get $\hat{\xi}_j$ if we get $\hat{\delta}_j$. Suppose $E[\xi Z] = 0$, then we can use Z as IV to estimate consistently.

Berry et al. (1995) (BLP henceforce) is more comprehensive with the same idea. They introduce random-coefficients logit model, which allows each individual to have different taste coefficients. Thus we can't directly use the inverse technique. The way is to use a Simulation Method of Moments (inner loop and outer loop). First guees a δ, match mkt share, then run a IV regression, then guess different δ and get an outer loop. They prove a contraction mapping.

Quan, T. W. and K. R. Williams (2017). Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail. SSRN Electronic Journal, 1–74

This paper addresses the local heterogeneity of consumer demand, and concludes that usual BLP method will overestimate the welfare increase of e-commerce because local stores have satisfied most local demands.

Challenges:

- Huge varieties of goods in fine-grid locations make scarce data in each category. Two possible solutions for this: 1) aggregating over locations; 2) just ignore the zero sales. Aggregating is unsatisfactory since the heterogeneity of local preference is our study focus. Ignoring will cause selection bias and cause overestimation of local heterogeneity.

- Mean utility level cannot be recovered solely by observable variables. They raise specific Converging in Distribution theorem to tackle this.

Erdem, T. and M. P. Keane (1996). Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. Marketing Science 15(1), 1–20

Erdem and Keane (1996) provides an example on how to estimate demand when quality is uncertain and consumer learns by Bayesian rule.

Motivation:

- Avoide ad hoc assumptions on how consumers utilizing experience information

- Estimate structural models to do policy analysis

Features:

- Estimate Dynamic Programming. Consumers make decision based on their perceptions, but econometricians only observe the consumption. So have to integrate over the distribution of perception errors $v_{ij}(t)$, which is is a Markov process. The integration part will be very complicated.

  - Provides a way to estimate Dynamic Programming model with simulation. Also see Keane and Wolpin (1994).

- Experience Shocks. This indicates that even after purchasing once, consumer is not certain about the quality of the good (or the preference of himself). This is mainly motivated by empirical fact: consumption behavior generally do not converge.

Empirical Findings:

- Bayesian learning very well fit the date.

11.2 BLP, 95 EMCA

The note on BLP draws deeply from Nevo (2000), the original paper Berry et al. (1995) is very worth reading.

Consumer Utility:

The consumer indirect utility is specified as:

$$u_{ijt} = \alpha_i (y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}$$

where i for consumer, j for products and t for time periods (markets); x_{jt} are the observable characteritics, ξ_{jt} is the unobservable product charateristics. Note that the quasilinear structure is not important, BLP beings from Cobb-Douglas form direct utility and end up with a log linear indirect utility form.

We also need to specify an outside good:

$$u_{i0t} = \alpha_i y_i + \xi_{0t} + \pi_0 D_i + \sigma_0 v_{i0} + \varepsilon_{i0t}$$

Consumer preference α_i and β_i is heterogeneous across individuals. They are specified by:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma v_i$$

where α and β are the common part; D_i is hte demographics of individual i, for which the distribution is known (for example from census data); v_i is the unobserved heterogeneous part, we assume it being multivariate normal. Π is a mapping from demographics to preference, and Σ stands for the covariance matrix of heterogeneous part.

The utility can be divided into two parts, the mean utility common for every consumer and the heterogeneous utility part:

$$u_{ijt} = \alpha_i y_i + \delta_{jt} (x_{jt'}, p_{jt}, \xi_{jt}; \theta_1) + \mu_{ijt} (x_{jt}, p_{jt}, v_i, D_i; \theta_2) + \varepsilon_{ijt}$$

where the δ_{jt} is the mean utility for product j in market t, and μ_{ijt} is the heterogeneous part. The parameters in δ_{jt} are called *linear* parameters, and parameters in μ_{ijt} are called *nonlinear* parameters.

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}, \quad u_{ijt} = [-p_{jt}, x_{jt}] (\Pi D_i + \Sigma v_i)$$

Market Share:

The market share can be computed by:

$$\begin{aligned} s_{jt}\left(x_{.t}, p_{\cdot t}, \delta_t; \theta_2\right) &= \int_{A_{jt}} dP^*(D, v, \varepsilon) \\ &= \int_{A_{jt}} dP^*_{\varepsilon}(\varepsilon) dP^*_{v}(\nu) d\widehat{P}^*_{D}(D) \end{aligned} \tag{11.1}$$

where A_{jt} is the set of all product characteristics that lead to choosing product j. Notice that the distribution of ε, ν and D are known to econometrician.

Estimation:

Notice that unobservable ξ_{jt} enters the market share non-linearly. But if we have δ_{jt}, then ξ_{jt} can be represented linearly by:

$$\xi_{jt} \equiv \delta_{jt}\left(S_{.t}; \theta_2\right) - \left(x_{jt}\beta + \alpha p_{jt}\right)$$

BLP shows that δ_{jt} can be calculated from observed market shares using contraction mapping:

$$\delta^{h+1}_{\cdot t} = \delta^{h}_{\cdot t} + \ln S_{\cdot t} - \ln S\left(p_{\cdot t}, x_{\cdot t}, \delta^{h}_{\cdot t}, P_{ns}; \theta_2\right)$$

Chapter 12

Entry and Exit

12.1 Berry, 92 EMCA

Berry, S. T. (1992). Estimation of a Model of Entry in the Airline Industry. Econometrica 60(4), 889

Highlights of the Paper:

- This paper studies the entry in airline industry. There are two issues to be focused: first is the simultaneity of of profits and market structure in oligopolistic market; second is the presence of a large number of heterogeneous potential entrants, which induces computational difficulty in estimation.
- This paper features **heterogeneity of firms**, and there is a large number of potential entrants. This paper shows that the SMM approach can solve this complicated equilibrium problem in a relatively easy way.

Model:

- The model is assumed to be **static**. At the beginning of each period, each firm takes the current network structure as given and decides whether to operate in a given city pair. The expost profit of a firm in a given market is determined by market level characteristics, its own airport presence, and of the number of competitors it faces in this market. Notice the profit is independent of the firm's profit in other markets.

- The profit function for firm k in market i is:

$$\pi_{ik}(s) = v_i(N(s)) + \phi_{ik} \tag{1}$$

where s is a given strategy file of all firms; $N(s)$ is the number of entering firms; and ϕ_{ik} is the index of profitability of firm k in market i.

  - [Firm Heterogeneity] We assume firm characteristics will only affect the fixed cost part, so the expost competition is symmetry. Thus $v_i(\cdot)$ only depends on the number of entering firms (up to scale), and the firm specific ϕ_{ik} only enters linearly. This assumption is relaxed in Ciliberto and Tamer (2009), which assumes a very general form of firm heterogeneity.
  - [Unique N^*] This separability of market level profit and firm level profit is essential. Given equation (1), all equilibriums feature a unique N^*. And this can be used as identification tool.
  - We further specify $v_i(N)$ and ϕ_{ik} parametrically:

$$v_i(N) = X_i\beta + h(\delta, N) + \rho u_{io} \tag{2}$$

    β, δ, ρ are parameters to be estimated. X_i are characteristics of market i, u_{io} are market characteristics observed by all firms but not econometricians. [1] This can be derived from a Cournot model with constant and identical marginal costs together with a constant-elasticity demand function.
  - Firm specific portion of profit is:

$$\phi_{ik} = Z_{ik}\alpha + \sigma u_{ik}$$

    Z_{ik} are firm level characteristics and u_{ik} unobservable to econometricians.
  - KZ: We have to assume some independence on unobservable characteristics for identification. But this is actually not very reasonable since in the firm side these characteristics are nothing special and thus may possess any kind of correlation.

Combine all these and we get:

$$\pi_{ik}(N) = X_i\beta - \delta\ln(N) + Z_{ik}\alpha + \rho u_{io} + \sigma u_{ik} \tag{4}$$

We denote the observed terms as $r_{ik}(N)$ and the unobserved terms as ε_{ik}. Notice that we can not identify ρ, σ separately since they are just scale parameters, and this discrete choice model is identified up to scale. This paper normalize σ to $\sqrt{1-\rho^2}$.

Identification:

[1] This is a complete information game.

- The identification is through the equilibrium number N^* of entering firms. For a given unobserved ε, the equilibrium entering firm number is smaller or equal than N iff:

$$(\varepsilon : \#\{\varepsilon_k : \varepsilon_k \geq -r_k(N+1)\} = j),\ j = 0, \ldots, N$$

  That is, no more than (but may be less than) N firms can make a profit if $N+1$ firms enter.

- Then suppose exact J firms can make profits in a $N+1$ equilibrium, the probability of this event is:

$$H_{J(N+1)} = \sum_{s \in S(J)} \int_{A_1(s,N+1)} \cdots \int_{A_k(s,N+1)} p(\varepsilon) d\varepsilon$$

  where s is a specific entering profile, and $S(J)$ is the set of entering profiles with J entering firms. For K potential entering firms, there are C_{N+1}^J many possible entering profiles. The area $A_k(s, N+1)$ is:

$$\begin{aligned} A_k(s, N+1) &= (-r_k(N+1), \infty),\ \text{if } s_k = 1 \\ &= (-\infty, -r_k(N+1)),\ \text{else} \end{aligned}$$

  That is, $A_k(s, N)$ is the region of ε_k that firm k can earn profit in a $(N+1)$-firm equilibrium if the profile s shows it enters. The probability $Pr(N^* \leq N)$ is:

$$Pr(N^* \leq N) = \sum_{J=0}^{N} H_{J(N+1)}$$

  And thus the probability of an N-firm equilibrium is:

$$Pr(N^* = N) = Pr(N^* \leq N) - Pr(N^* \leq (N-1))$$

  We can then apply MLE to maximize this probability.

- Notice that although in principle we can estimate this in MLE as long as we have enough variation in X_i, the computation of $Pr(N^* = N)$ is very difficult. This paper tackles this difficulty with two approaches: 1) they put additional restrictions over the parameters; 2) they use simulation methods to simplify.

Computation Issues:

- Although in principle we can identify the parameter using MLE using above expression, but the computation burden is huge. As we have pointed out, the number of possible combinations of potential entrants and actual entrants are C_n^k, which grows very fast. To solve this problem, this paper takes two approaches.

- Put more restrictions over the parameters
- Use SMM

We focus on the second approach.

- [SMM] First we generate a series of $\hat{u}_i$, then given θ, we can calculate an estimate $\hat{N}$ for each market, with the property:

$$E_{\hat{u}}[\hat{N}(W_i, \theta, \hat{u})] = E[N^*|W_i, \theta]$$

So we can write the moment condition as:

$$v_{io}(N_i^*, W_i, \theta) \equiv N_i^* - E_{\hat{u}}[\hat{N}(W_i, \theta, \hat{u})]$$

Estimating this moment condition will give us consistent θ.

Empirical Results:

- Firm observable heterogeneities, such as airport presence, play an important role in determining airline profitability.
- Profits decline rapidly in the number of entering firms, consistent with Bresnahan and Reiss (1990).

Pros:

-

Cons:

- The unobserved term u_{ik} and u_{io} are assumed to be i.i.d. across both firms and markets. Especially, u_{ik} are more likely to be correlated across markets for a given firm.
- The heterogeneity is assumed to be a very specific term, which only affects the “fixed cost” of firms. Conditional on firm entering, the subsequent subgame is symmetric. This restriction is quite strong.

12.2 Seim, 06 RAND

Seim, K. (2006, September). An empirical model of firm entry with endogenous product-type choices. The RAND Journal of Economics 37(3), 619–640

Highlights of the Paper:

- By introducing private information and using Bayesian Nash Equilibria, the strategy of each firm is a mapping from its own type c_i to a strategy s_i, which does not directly depend on other firms' strategies. This ensures the uniqueness of equilibrium in the specific setting of this paper, but uniqueness is not guaranteed in general.

Model:

-

Empirical Results:

-

Pros:

-

Cons:

- Charlie Murry points out that the result is somewhat directly being assumed. The potential number of entrants $\mathscr{F}$ is being assumed, but if we assume this and we know the actual entrant number $\mathscr{E}$, isn't the probability of entrant directly derived?

12.3 Davis, 06 JoE

Davis, P. (2006, September). Estimation of quantity games in the presence of indivisibilities and heterogeneous firms. Journal of Econometrics 134(1), 187–214

Highlights:

- This paper is a natural generalization of Berry (1992). Berry guarantees the uniqueness of entering firms using assumption of "separable profit function" and "symmetric second stage competition". This paper relaxes the assumptions by supermodularity conditions.
- Under his assumptions, this paper guarantees: 1) existence of PSNE; 2) the uniqueness of market index (number of entering firms in Berry's model) within PSNE.

Model:

-

Empirical Results:

-

Pros:

-

Cons:

-

12.4 Jia, 08 EMCA

Jia, P. (2008). What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry. Econometrica 76(6), 1263–1316

Highlights of the Paper:

- Jia's paper does not investigate the crowdness of stores in one market. It focuses on the distribution of stores across the country. In Jia's market segmentation, each brand will put at most one store in each market. So the whole story is about how to allocate stores in different markets, instead of how many stores in one market.

12.5 Tamer, 03 RES

Tamer, E. (2003). Incomplete Simultaneous Discrete Response Model with Multiple Equilibria. Review of Economic Studies 70(1), 147–165

Highlights:

- This paper shows that "games with multiple equilibria can be estimated without imposing coherency conditions if the model is used only to inform upper and lower bounds on the probabilities of outcomes rather than assigning exact probabilities to each potential outcome." (Davis, 2006)
- This paper differentiates **incoherent model** with **incomplete model**. *Incoherency* comes from the not well defined optimization problem of the agent; while *incomplete* econometric models means that the mapping from (x, u) to y is not a function but a correspondence.
- The paper also shows, the parameter of an incomplete model can be **point identified**, as long as we have at least one exogenous variable with unbounded range. This is called identification at infinity.

Identification:

- For a model like in Bresnahan and Reiss (1990):

$$\begin{aligned} y_1 &= x_1\beta_1 + y_2\Delta_1 + u_1 \\ y_2 &= x_2\beta_2 + y_1\Delta_2 + u_2 \end{aligned}$$

- If there is one exogenous variable (exclusively) on x_1 or x_2 that can take value at infinity, then the parameter is point identified. The idea is simple, say x_{1k} takes $-\infty$ and β_{1k} is positive, then firm 1 will not enter the market for sure regardless of (u_1, u_2). Thus all other parameters can be point identified using this subsample.
- [β Identification] Take x_{1k} to $-\infty$:

$$\Pr[(0,0)|x_1, x_2^*] = \Pr\left[u_1 \le -x_1\beta_1, u_2 \le -x_2^*\beta_2\right] \neq \Pr\left[u_1 \le -x_1 b_1, u_2 \le -x_2^* b_2\right] \tag{12.1}$$

$$\tag{12.2}$$

the inequality is from:

$$\begin{aligned} \Pr\left[u_1 \le -x_1\beta_1, u_2 \le -x_2^*\beta_2\right] &\simeq \Pr\left[u_2 \le -x_2^*\beta_2\right] \neq \Pr\left[u_2 \le -x_2^* b_2\right] \\ &\simeq \Pr\left[u_1 \le -x_1 b_1, u_2 \le -x_2^* b_2\right] \end{aligned}$$

Thus β_2 is identified. Given β_2, we can identify β_1 since x_1 is full rank.

- [Δ Identification] The identification of Δ is similar to that of θ. The difference is that we now use $(y_1, y_2) = (1,1)$ instead of $(0,0)$.

$$\Pr[(1,1)|x_1, x_2^*] = \Pr(u_1 > -x_1\beta_1 - \Delta_1, u_2 > -x_2\beta_2 - \Delta_2) \neq \Pr(u_2 > -x_2\beta_2 - d_2) \tag{12.3}$$

- Only observations of $(0,0)$ and $(1,1)$ are necessary for identification. The main point to use $(0,1)$ and $(1,0)$ is to increase efficiency with finite data.

Semi-Parametric Estimation:

- Define the population frequency of $(0,1)$ to be $H(\cdot)$:

$$H(x) = \Pr[(0,1)|x]$$

and the quasi-likelihood can be defined as:

$$\begin{aligned} L\left(b;\hat{H}_{tn}\right) = \sum_{i=1}^{n} & [y_{i1}y_{i2}\log\left(P_1\left(x_i,b\right)\right) \\ & + \left(1-y_{i1}\right)\left(1-y_{i2}\right)\log\left(P_2\left(x_i,b\right)\right) + \left(1-y_{i1}\right)y_{i2}\log\left(\hat{H}_{tn}\left(x_i,b\right)\right) \\ & + y_{i1}\left(1-y_{i2}\right)\log\left(1-P_1\left(x_i,b\right)-P_2\left(x_i,b\right)-\hat{H}_{tn}\left(x_i,b\right)\right)] \end{aligned}$$

where P_1, P_2 can analytically be derived and H is non-parametrically estimated.

- The SML β is consistent and normally distributed (in BR case).

12.6 Ciliberto and Tamer, 09 RES

Ciliberto, F. and E. Tamer (2009). Market Structure and Multiple Equilibria in Airline Markets. Econometrica 77(6), 1791–1828

Highlights of the Paper:

- This paper has two contributions to entry game estimation literature. First, it allows very general form of firm heterogeneity; second, by adopting partial identification it makes no assumption on equilibrium selection.

- [Firm Heterogeneity] Recall that in Berry (1992), the only heterogeneity is a heterogeneous "fixed cost", which enters linearly. If we allow more general form of heterogeneity, then even the number of entrants is not unique.

- [Multiple Equilibria] The estimation procedure allows mixed strategy. But this will make the equilibrium solving process more difficult. The model **does not specify or estimate any equilibrium selection rule** ($\Pr(y|\varepsilon, X)$ in the paper). In principle this can be estimated, either non-parametrically (as in Tamer (2003)) or as a function of ε and X (as in Berry and Tamer (2006)). This makes the identified set not sharp, but the computation much easier.

Model:

- The profit function for firm i in market m is:

$$\pi_{im} = S'_m \alpha_i + Z'_{im}\beta_i + W'_{im}\gamma_i + \sum_{j\neq i} \delta^i_j y_{jm} + \sum_{j\neq i} Z'_{jm}\phi^i_j y_{im} + \varepsilon_{im}$$

where S_m is market characteristics; W_{im} is the firm characteristics that will only affect its own profit in market m, such as cost variables; Z_{im} is the firm characteristics that may affect both own and others profit. $\sum_{j\neq i} \delta^i_j y_{jm}$ captures the effect of other firm's entering, and $\sum_{j\neq i} Z'_{jm}\phi^i_j y_{im}$ further captures the effect of entering firms' characteristics.

  - Notice that the impact of other firms can be directly dependent on firm identity. For a specific firm i, having j or having k in the market may impact i's profit differently, even if j and k has same characteristics on this market.

- [Main Idea] Conditional on observables, although we cannot calculate *the probability* of a entering profile y if multiple equilibria exist, but we can get the *bound* of probabilities.

$$\begin{aligned} \underline{\Pr}(y) &= \Pr(\vec{\varepsilon}| \text{ y is the unique equilibrium}) \\ \overline{\Pr}(y) &= \Pr(\vec{\varepsilon}| \text{ y is among the equilibria}) \end{aligned} \tag{12.4}$$

- [Identification] The identified set Θ_I contains all θ's s.t. the *all moment inequalities* hold. Point identified at infinite exogenous variable, see Tamer (2003). In practical estimation, variation in excluded exogenous variables (like the airport presence or cost variables in this paper) help shrink the set Θ_I.

Estimation:

- The estimation is based on the moment inequalities:

$$\mathbf{H}_1(\boldsymbol{\theta}, \mathbf{X}) \equiv \begin{bmatrix} H_1^1(\boldsymbol{\theta}, X) \\ \vdots \\ H_1^{2K}(\boldsymbol{\theta}, X) \end{bmatrix} \leq \begin{bmatrix} \Pr(\mathbf{y}_1 | X) \\ \vdots \\ \Pr(\mathbf{y}_{2^K} | X) \end{bmatrix} \leq \begin{bmatrix} H_2^1(\boldsymbol{\theta}, X) \\ \vdots \\ H_2^{2^K}(\boldsymbol{\theta}, X) \end{bmatrix} \equiv \mathbf{H}_2(\boldsymbol{\theta}, \mathbf{X}) \tag{12.5}$$

  The objection function is:

$$Q(\boldsymbol{\theta}) = \int \left[\|(P(\mathbf{X}) - H_1(\mathbf{X}, \boldsymbol{\theta})) - \| + \left\| (P(\mathbf{X}) - H_2(\mathbf{X}, \boldsymbol{\theta}))_+ \right\| \right] dF_x$$

  Notice that $\mathbf{H}_1$ and $\mathbf{H}_2$ cannot be analytically derived in a complicated game, and have to be estimated by simulation.

- [Θ_I or θ_I] When doing infernece, the interest is either the identified set Θ_I or the true parameter $\theta_I \in \Theta_I$. Note that there is one true θ_I even if Θ_I is only set identified (e.g. when the unbounded exclusion variable x is not available).

Empirics:

- To be added: the use of hub airports
- To be added: the marginal effect table

Cons:

- The identified set is not sharp. What is a sharp identified set?
- Does the effect of other firms δ have to be negative?

12.7 Grieco, 14 RAND

Grieco, P. L. E. (2014, June). Discrete games with flexible information structures: an application to local grocery markets. The RAND Journal of Economics 45(2), 303–340

Highlights:

- This paper integrate complete information entry game and incomplete information entry game in one model. The paper also tests the two polar case of complete information and incomplete information.

- The intuition for the identification relies on instrument variable that affects private information but not public information. In the private information game, player 1's decision will not depend on private information x_2, because player 1 can only take into account public informations and make a guess on player 2's decision. Contrary to this, private information x_2 will affect player 1's decision in a complete information game if the covariates support is rich enough. Thus if we can find some instruments affecting private information (x_1, x_2), we can identify complete and incomplete information.

Model:

- The model is a two firm entry problem with binary decisions:

$$\pi_i\left(y_i, y_{-i}; x, \theta\right) = \begin{cases} x_i\theta_{i\mu} + y_{-i}x_i\theta_{i\delta} + \epsilon_i + v_i & \text{if } y_i = 1 \\ 0 & \text{if } y_i = 0 \end{cases}$$

where the ε_i represents public shock, and μ_i is private shock. The public shocks $(\varepsilon_1, \varepsilon_2)$ follows bivariate normal, with common variance component σ_ε^2 and correlation coefficient ρ. Private shocks are independent and have variance $\sigma_\mu^2 \equiv 1 - \sigma_\varepsilon^2$. σ_ε^2 controls relative importance of public shocks.

- The equilibrium of this model is:

$$\chi_i^b\left(\chi_j, \epsilon, x; \theta\right) = -\left(x_i\theta_{i\mu} + \varrho_j\left(\chi_j, \epsilon; \theta\right)x_i\theta_{i\delta} + \epsilon_i\right)$$

where the left is the threshold of player i to enter, and ϱ_j is the expected probability of j to enter, conditional on j's threshold. Any joint set $\chi = (\chi_1, \chi_2)$ satisfies the above equation will be an equilibrium.

- [Multiplicity] Note that even with incomplete shock only, there may exist multiple equilibria. Though each player will play a threshold strategy, the values of thresholds are coordinated. Note the difference with Seim (2006).

- Grieco finds that the area of multiplicity shrinks as the information environment becomes more and more incomplete. See their Fig 1.

Identification:

- [Only Private Shock Case] The probability of observing outcome y given a selected equilibrium e is:

$$\tilde{\Psi}(y|x, \theta, e) = \prod_{i=1}^{2} \varrho_i^e(x, \theta)^{1[y_i=1]}\left(1 - \varrho_i^e(x, \theta)\right)^{1[y_i=0]}$$

Unconditional on equilibrium, it can be represented as:

$$\Psi(y|x,\theta,\lambda) = \sum_{e\in\delta(x,\theta)} \lambda^e(x,\theta)\tilde{\Psi}(y|x,\theta,e)$$

where λ^e is the equilibrium selection rule and can be jointly estimated. See Tamer (2003) semi-parametric estimation routine. The sharp identified set is:

$$\Theta_I = \begin{Bmatrix} \theta\in\Theta : \forall y,x,\exists\tilde{\lambda}\in[0,1]^{\overline{E}} \text{ s.t. } P(y|x) = \Psi(y|x,\theta,\tilde{\lambda}) \\ \sum_{e\in\mathscr{E}(x;\theta)} \tilde{\lambda}^e = 1 \end{Bmatrix}$$

How we are going to estimate this? are the set convex or something?

Inference

- This paper uses a sieve based weighted Bootstrap to do inference. But Karl says there is something wrong with this approach. Not sure about this.

12.8 Holmes, 11 EMCA

Estimation

Disagregate Inequalities v.s. Aggregate Inequalities

Why not using disaggregate inequalities which will generate more moment inequalities? Because there is measurement error. Suppose the model is:

$$y_a \geq \theta$$

but the LHS is is observed with error:

$$\tilde{y}_a = y_a + \eta_a$$

If we pick θ to minimize the disaggregated inequalities, then we are looking for θ that satisfies:

$$\left\{\theta|\theta \leq \min_a\{y_a+\eta_a\}\right\}$$

and some very extreme η_a will drive the results.

12.9 Lee and Pakes, 09 EL

There is multiplicity problem in counterfactual analysis of entry game. This paper discusses two possible ways of dealing multiple equilibria.

Enumerating Equilibria

The first approach just enumerate all possible limiting equilibria, and then compare them. The equilibria are limiting in the sense that all firms know the expected value of the cost shock. The good news is, there are only a handful of possible equilibria in their case.

Equilibria Selection by Learning

Based on previous enumeration results, we can apply learning process to see which equilibrium is more possible. There are two possible learning processes:

- [Best Response Dynamics] Firm believes its competitors will play the same strategy as they did in the prior period, and thus each firm will play its best response to last periods' play.

- [Fictitious Play] Each firm believes that the next play of its competitors will be a random draw from the set of tuples of plays observed since the regime change, and chooses it's optimal strategy accordingly. (KZ: the optimal should be mix?)

To get the distribution of equilibria, they compute 1000 runs for each learning process. Each run is stopped when they converged to a single allocation profile, where convergence is defined as having remained in the same state for 50 iterations.

12.10 Wollman, 18 AER

Identification

The identification of the second stage is a standard BLP, I will focus on the identification of the first step about the fixed costs.

The inequalities based on rationality condition is:

$$\begin{aligned}&\Delta\pi\left(J_{f,t}, J_{f,t}\backslash j, J_{-f,t}, z_t, w_t\right)\\&\quad + x_j'\left(\{j \notin J_{f,t-1}\} - \lambda\{j \in J_{f,t-1}\}\right)\mathscr{E}\left[\theta_{f,x_j,t}|\mathscr{J}_{f,t}\right] \geq 0\end{aligned} \tag{6}$$

$$\begin{aligned}&\Delta\pi\left(J_{f,t},J_{f,t}\cup j,J_{-f,t},z_t,w_t\right)\\&-x_j'\left(\{j\notin J_{f,t-1}\}-\lambda\{j\in J_{f,t-1}\}\right)\mathscr{E}\left[\theta_{fx_j,t}|\mathscr{I}_{f,t}\right]\geq 0\end{aligned}\tag{7}$$

Note that $\mathscr{E}[\theta_{f,x_j,t}|\mathscr{F}_{f,t}]$ is unknown to econometricians, thus we need to parameterize it and convert the inequalities to some observable moments. By parameterization:

$$\theta_f=\left[\theta_0+\theta_0^{Big3}+\theta_0^{Jpn},\theta_g+\theta_g^{Big3},\theta_{cabover}+\theta_{cabover}^{Jpn},\theta_{compact},\theta_{long}\right]$$

and the suck cost parameter $\theta_{f,x_j,t}\equiv\theta_f+\nu_{2,t}$. Notice that the structural error $\nu_{2,t}$ is a vector, with each element indexing a characteristic error. To rationalize data, we need to introduce structural error $\nu_{2,t}$ and expectation error $\nu_{1,t}$.

- ν_2 is known to the firms when making entry decision, but unknown to econometricians
- ν_1 is unknown to firms when making entry decision

The actual moment inequalities for estimation are:

$$\begin{aligned}&\tfrac{1}{XTF}\textstyle\sum_{x_j}\sum_t\sum_f h^i_{f,x,t}\{j\in J_{f,t}\}\\&\times\left[\Delta\hat{\pi}\left(J_{f,t},J_{f,t}\backslash j,J_{-f,t},z_t,w_t\right)+x_j'\theta_f\left(\{j\notin J_{f,t-1}\}-\lambda\{j\in J_{f,t-1}\}\right)\right]\geq 0\end{aligned}\tag{8}$$

$$\begin{aligned}&\tfrac{1}{XTF}\textstyle\sum_{x_j}\sum_t\sum_f h^i_{f,x_j,t}\{j\notin J_{f,t}\}\\&\times\left[\Delta\hat{\pi}\left(J_{f,t},J_{f,t}\cup j,J_{-f,t},z_t,w_t\right)-x_j'\theta_f\left(\{j\notin J_{f,t-1}\}-\lambda\{j\in J_{f,t-1}\}\right)\right]\geq 0\end{aligned}\tag{9}$$

In appendix, the empirical moment is shown to converge in probability to:

$$\begin{aligned}&\xrightarrow{p}E_{x_j,t,f}\left[h^i_{f,x_j,t}\{j\in J_{f,t}\}\Delta\pi\left(J_{f,t},J_{f,t}\backslash j,J_{-f,t},x_t,w_t\right)\right]\\&+E_{x_j,t,f}\left[h^i_{f,x_j,t}\{j\in J_{f,t}\}\left[x_j'\left(\{j\notin J_{f,t-1}\}-\lambda\{j\in J_{f,t-1}\}\right)\mathscr{E}\left[\theta_{f,x_j,t}|\mathscr{F}_{f,t}\right]\right]\right]\\&+E_{x_j,t,f}\left[h^i_{f,x_j,t}\{j\in J_{f,t}\}\left(\{j\notin J_{f,t-1}\}-\lambda\{j\in J_{f,t-1}\}\right)\nu_{1,f,t,J_{f,t},J_{f,t}\backslash j,J_{f,t-1}}\right]\\&-CE_{x_j}\left[x_j'h^i_{f,x_j}E_t\left[h^i_{x_j,t}\nu_{2,t}\right]\right]\geq 0\end{aligned}$$

The first two lines larger than 0 by rationality assumption, the third line is 0 in expectation since ν_1 is mean independent of agents' information set. The forth line is tricky, and we need to pick h^i to make it zero in expectation.

"Note that $h^i_{x_j}$ selects observations based on product characteristics. For example, if the i th moment considers conventional vehicles only, then $h^i_{x_j}=1$ when the cabover , compact , and long dummy variables in x_j equal 0, and $h^i_{x_j}=0$ otherwise. Further, $h^i_{x_j,t}$ depends on the demand shifters specific to the subset of product types and action under consideration. For example, the construction industry accounts for the largest proportion of medium GWR vehicles. Hence, if the i th moment considers the entry of medium GWR vehicle buyers, then $h^i_{x_j,t} = 1$ for t in which the proportion of prospective buyers from the construction industry is particularly high. Finally, $\tilde{h}^i_{f,x_j,t} = 1$ selects observations such that either $j \in J_{f,t-1}$ or $j \notin J_{f,t-1}$. For example, if the ith moment considers entry, then $\tilde{h}^i_{f,x_j,t} = 1$ when $j \notin J_{f,t-1}$. Note that C is a constant equal to 1 if $\tilde{h}^i_{f,x_j,t}$ selects $j \notin J_{f,t-1}$ and equal to $-\lambda$ otherwise, as described below."

Things to be noted here:

1. $\mathscr{E}$ represents agent's subjective expectation, and E is the expectation by actual DGP.
2. The unobservable $\mathscr{E}[\theta_{f,x_j,t}]$ has been substituted by parameter θ_f. We have to take care of ν_2 because it may induce selection problem.
3. $h^i_{f,x_j,t}\{j \notin J_{f,t}\}$ is used as instruments and to control the endogenous problem induced by ν_2.

Instrument Variables

As in many PPHI application papers, the IVs here are just a set of indicator functions to pick selected observations.

The idea is to use demand side shifters to pick periods when certain types of products will surely be introduced, and to pick periods when certain types will surely be retired. For these instruments to be valid, it requires that there is no selection of ν_2 in the subset of observations picked by instruments h_i. Why such h_i will not induce the selection of ν_2? This is like a Tamer (2003) argument. If there are some periods that the demand for product j is super high, then someone will introduce the product regardless of ν_2.

The key is that the structural error term ν_2 is characteristic-year specific rather than firm-product-year specific, so while the inequalities vary at the firm-product-year level, the structural error term does not. At each time period, $\nu_{2,t}$ is a vector with 6 elements. The same ν_2 applies to all firm-products as long as their characteristics are the same.

Chapter 13

Dynamic Decisions

13.1 Rust, 87 EMCA

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. Econometrica 55(5), 999

Highlight of the Paper:

- This is the seminal paper to estimate a dynamic model. A very clever trick is separating EV_θ with a Bellman function, which greatly reduces the dimension since ε is not an argument of EV.
- But the computation burden is still heavy, because for every θ we need to solve a contraction mapping. Hotz and Miller (1993) eases the computation using SMM.

Structural Estimation without Closed-form Solutions:

- The decision maker has the following value function:

$$V_\theta(x_t, \varepsilon_t) = \sup_{\Pi} E\Big\{\sum_{j=t}^{\infty} \beta^{j-t}[u(x_j, f_j, \theta_1) + \varepsilon_j(f_j)]|x_t, \varepsilon_t, \theta_2, \theta_3\Big\}$$

where $f_t(x_t, \varepsilon_t, \theta)$ is the optimal control function. The transition rule is:

$$dp\{x_{t+1}, \varepsilon_{t+1}, \ldots, x_{t+N}, \varepsilon_{t+N}\} = \prod_{i+t}^{N-1} p(x_{i+1}, \varepsilon_{i+1}|x_i, \varepsilon_i, f_i(x_i, \varepsilon_i), \theta_2, \theta_3)$$

This infinite horizon problem can be rewritten as a Bellman function with stationary decision rule:

$$V_\theta(x_t, \varepsilon_t) = \max_{i \in C(x_i)} [u(x_i, i, \theta_1) + \varepsilon_t(i) + \beta EV_\theta(x_t, \varepsilon_t, i)]$$

where

$$EV_\theta(x_t, \varepsilon_t, i) = \int_y \int_\eta V_\theta(y, \eta) p(dy, d\eta | x_t, \varepsilon_t, i, \theta_2, \theta_3)$$

One property (difficulty) of this model is that distribution of ε_{t+1} is conditional on i and ε_t. So EV_θ is a nonlinear function of ε_t, which would induce huge dimensionality problem in value function iteration.

- This model is actually a simple version of Cooper's general durable consumption problem. The key is that $\{x_t\}$ follows different transition rule after different action i. This problem is actually simpler because the only decision variable is i.

- [Conditional Independence Assumption] To make this problem computationally tractable, we assume:

$$p(x_{t_1}, \varepsilon_{t+1} | x_t, \varepsilon_t, i, \theta_2, \theta_3) = q(\varepsilon_{t+1} | x_{t+1}, \theta_2) p(x_{t+1} | x_t, i, \theta_3)$$

This assumption says, the transition of (x_t, ε_t) is separable. Probability of ε_{t+1} only directly depends on x_{t+1}, and the transition of x_{t+1} is independent of current ε_t. That's to say, ε_t is like a temporary disturbance on x_t.

- [Below Important] The power of above assumption is that we now can solve EV_θ in a contraction mapping T_θ (the proof is not trivial):

$$EV_\theta(x, i) = \int_y G(u(y, \theta_1) + \beta EV_\theta(y) | y, \theta_2) \cdot p(dy | x, i, \theta_3)$$

where $V_\theta(x)$ actually is $\max\{V_\theta(x, 0), V_\theta(x, 1)\}$, and

$$G(u(x, \theta_1) + \beta EV_\theta(x) | x, \theta_2) = \int_\varepsilon \max_{j \in C(x)} [u(x, j, \theta_1) + \beta EV_\theta(x, j)] q(d\varepsilon | x, \theta_2)$$

And the conditional probability of choosing choice i is given by:

$$P(i | x, \theta) = G_i(u(x, \theta_1) + \beta EV_\theta(x) | x, \theta_2)$$

Now we can define the likelihood function: (directly from conditional independence assumption)

$$\mathscr{L}^f(x_1, i_1, \ldots, x_T, i_T | x_0, i_0, \theta) = \prod_{t=1}^{T} P(i_t | x_t, \theta) p(x_t | x_{t-1}, i_{t-1}, \theta_3)$$

- The transition rule parameter θ_3 can be separately estimated.

- [Nested Fixed Point Algorithm] An inner fixed point algorithm solves EV_θ for each value of θ; an outer algorithm captures θ that maximizes the likelihood.
  - To estimate EV_θ, we need to further assume that ε's are i.i.d. distributed across choices and time according to T1EV. With this assumption, we have:

$$\begin{aligned} EV_\theta(x,i) &\equiv E_{y,\varepsilon|x,i}\Big\{\max_{j=0,1}[u(y,\varepsilon;\theta)+\beta EV(y,j;\theta)]\Big\} \\ &= E_{y|x,i}\log\Big\{\sum_{j=0,1}\exp[u(y,\varepsilon;\theta)+\beta EV(y,j;\theta)]\Big\} \\ &= \int_y log\Big\{\sum_{j=0,1}\exp[u(y,j;\theta)+\beta EV(y,j;\theta)]\Big\}p(dy|x,i) \end{aligned}$$

    The second equality uses the closed form of maximization over expectation of T1EV. [1]
  - Once $EV_\theta(x,i)$ has been calculated using value function iteration, the choice probability $p(i_t|x_t)$ can be computed as:
$$\frac{\exp(E_\varepsilon u(x_t,\varepsilon,i_t;\theta)+\beta EV(x,i;\theta))}{\sum_{i=0,1}\exp(E_\varepsilon u(x_t,\varepsilon,i_t;\theta)+\beta EV(x,i;\theta))}$$
  - Once we get $EV_\theta(x,i)$, then we can get $P(i_t|x_T,\theta)$. And since the transition rule $p(x_t|x_{t-1},i_{t-1},\theta_3)$ can be separately estimated. We now have collected all components of the likelihood function. Loop over θ to maximize the likelihood function.

Empirical Results:

-

Pros:

- The *Conditional Independence* assumption of ε is very important. With this assumption, the ε_T becomes a perturbation of x_t and independent of everything else. So although ε is still a state variable of V, it is no longer a state variable of EV. Then we can first solve $EV(x)$ by Bellman iteration.

Cons:

- The inner fixed point loop is computationally burdensome. And the complexity will significantly grow if we add more state variables. Hotz and Miller (1993) uses an inversion function to avoid the Bellman iteration.

[1] See Chiong, Galichon, and Shum (2016) for details.

13.2 Hotz and Miller, 93 RES

Hotz, V. J. and R. A. Miller (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. The Review of Economic Studies 60(3), 497

Highlights of the Paper:

- The main idea of this paper is to **ease the computation burden** of calculating the value function in Rust (1987). HM shows that value function can be represented in terms of choice probabilities, transition probabilities and expected per period payoffs. This hinges on the **invertibility** similar in BLP.

Model:

- The model uses a similar framework as Rust (1987). The agent is solving a single agent dynamic programming problem, the choice set is from $1, \ldots, J$. Denote the history to be H_t, which is a collection of period outcomes $b_t \in \mathbb{R}$, where b_t is affected by choice d_t. The history is generated according to the transition probabilities:

$$F_j(H_{t+1}|H_t)$$

  - H_t is the state variable here, corresponding to x_t in Rust paper. The transition rule is exogenous conditional on decision.

In each period t, there is a current utility u_{tj} associated with each choice j. Let $u_t^*(H_t) = E(u_{tj}|H_t)$.

$$u_{tj} = u_j^*(H_t) + \varepsilon_{tj}$$

  - Notice that by construction ε_t conditional on history H_t is independent of other $\varepsilon_{t'}$. The *conditional independence* assumption in Rust (1987) is implicitly assumed. The state variable x_t in Rust paper is included in H_t. Also, ε_{tj} is known to the agent. The distribution of ε_t is $G(\varepsilon_t|H_t)$.

The agent chooses a sequence $\{d_t\}_t \in T$ to maximize summed discounted utility.

Denote $p_k(H_t)$ to be the probability of choosing action k at history, and $p(H_t)$ be a $(J-1)$ vector of conditional choice probability function.

- Due to the conditional independence condition, we can write the probability of choosing choice 1:

$$\begin{aligned}
p_1(H_t) &= E(d_{t_1} = 1|H_t) \\
&= \int_{\varepsilon_1=-\infty}^{\infty} \int_{\varepsilon_2=-\infty}^{\varepsilon_1+u_{11}^*-u_{12}^*+v_{t1}-v_{t2}} \cdots \int_{\varepsilon_J=-\infty}^{\varepsilon_1+u_{11}^*-u_{1J}^*+v_{t1}-v_{tJ}} dG(\varepsilon_1,\ldots,\varepsilon_J|H_t) \\
&= \int_{-\infty}^{\infty} G_1(\varepsilon_1,[\varepsilon_1+u_{11}^*-u_{12}^*+v_{t1}-v_{t2}],\ldots,\varepsilon_1+u_{11}^*-u_{1J}^*+v_{t1}-v_{tJ}|H_t)d\varepsilon_1 \\
&\equiv Q_1(\tilde{v},H_t)
\end{aligned}$$

Similarly we can write $p_j(H_t)$ for each j. Then if we denote $\tilde{v} = (v_{t1} - v_{tJ}, \ldots, v_{t,J-1} - v_{t,J}) \equiv \tilde{v}(H_t)$. We can write:

$$p(H_t) = Q(\tilde{v}(H_t), H_t)$$

  – Notice that any $v_{t,j} - v(t,j')$ can be represented by two elements of $\tilde{v}$. $\tilde{v}$ just says that the scale of v is unidentified.

The key idea of estimation is to express $\tilde{v}(H_t)$ as a function of $p(H_t)$. This requires $Q(\tilde{v}, H_t)$ being **invertible** in $\tilde{v}$ for every H_t.

- KZ Guess: If $\tilde{v}$ can be expressed as function of $p(H_t)$, then we can use empirical $\hat{p}(H_t)$ to get empirical $\hat{v}$. If we further assume a functional form of ε distribution to be F1EV, then we can get an estimated choice probability $\bar{p}$. Then we can use GMM to match $\hat{p}$ and $\bar{p}$.

Empirical Results:

-

Pros:

-

Cons:

-

13.3 Bajari-Benkard-Levin, 07 EMCA

Bajari, P., C. L. Benkard, and J. Levin (2007, September). Estimating Dynamic Models of Imperfect Competition. Econometrica 75(5), 1331–1370

Question of the Paper:

- This paper raises a two-stage algorithm for estimating dynamic games under the assumption of Markov Perfect Equilibria. In first step, policy function and state variable transition rule are estimated. In second step, remaining structural parameters are estimated.

First Stage Estimation:

-

Second Stage Estimation:

-

Pros:

-

Cons:

-

Chapter 14

Asymmetric Information

14.1 Chiappori-Salanie, 00 JPE

Chiappori, P. A. and B. Salanie (2000, February). Testing for Asymmetric Information in Insurance Markets. Journal of Political Economy 108(1), 56–78

This is a first paper empirically testing *asymmetric information.* The **pair probits methodology** is important.

Identification:

- Suppose a pair probits system:

$$y_i = 1(X_i\beta + \varepsilon_i > 0)$$

$$z_i = 1(X_i\gamma + \eta_i > 0)$$

  where y_i is the probability of choosing high coverage, and z_i is the probability of accident. If there is no asymmetric information, then we should have ε_i and η_i independent. **If there is positive correlation, indicates negative selection; if negative correlation, indicates advantageous selection.**

- We first estimate these two probits separately. Then it provides a way of testing independence.

$$W = \frac{\left(\sum_{i=1}^{n} \omega_i \hat{\varepsilon}_i \hat{\eta}_i\right)^2}{\sum_{i=1}^{n} \omega_i^2 \hat{\varepsilon}_i^2 \hat{\eta}_i^2}$$

W should follow a $\chi^2(1)$ distribution. [1] Where $\hat{\varepsilon}_i$ and $\hat{\eta}_i$ are generalized residuals.

$$\hat{\varepsilon}_i = E(\varepsilon_i|y_i) = \frac{\phi(X_i\beta)}{\Phi(X_i\beta)}y_i - (1-y_i)\frac{\phi(X_i\beta)}{\Phi(-X_i\beta)}$$

- We can modify the test to get estimation of correlation between ε_i and η_i.

Empirical Results:

- Data: Car Insurance in France.
- Conditioning on the subset of young drivers, no significant evidence for adverse selection.
  - This paper strikes back Puelz and Snow (1994) in specific, which is criticized to suffer omitted observable variable problem.

Pros:

- Data requirement is low. Only need the individual choice of coverage, and individual accident probability.
  - Do not need premium information. Do not need exogenous price variation, which is needed in Einav, Finkelstein, and Cullen (2010).

Cons:

- Cannot disentangle moral hazard, which may induce correlation as well.
  - KZ: an ideal data will have z_i data before and after the exogenous insurance coverage change. Then we can directly test moral hazard.
- Only uses the coverage information, many other information about insurance (like premium, coinsurance rate) are not used.

14.2 Cardon-Hendel, 01 RAND

Cardon, J. H. and I. Hendel (2001). Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey. The RAND Journal of Economics 32(3), 408

[1] This is why we use a probit system.

This paper sets up a two stage model to investigate the asymmetric information in health insurance market. The adverse selection mainly happens in the first stage. The do not find significant evidence for adverse selection.

Question of the Paper:

- Use a two-stage model to investigate the asymmetric information in health insurance market.

Model:

- Agent do a two stage model to decide consumption of health insurance in first stage, and health care consumption in second stage.
- In second stage, given realized health state s_i and insurance policy j, each agent i decides how much health care to consume:

$$U_{ij}^*(s_i) = U^*(y_i, s_i, Z_j) = \max_{x_i} U(m_i, h_i)$$
$$s.t.\ m_i + C_j(x_i) = y_i - p_j$$

  where m_i is the consumption good and h_i is the health consumption. $h_i = x_i + s_i$, i.e. health care and health state are perfect substitution. y_i is income, p_j is premium of policy j, $C_j(x_i)$ is out of pocket money for consumption x_i with policy j.
- In first stage, the

Empirical Results:

-

Pros:

-

Cons:

-

14.3 Einav-Finkelstein-Cullen, 10 QJE

Einav, L., A. Finkelstein, and M. R. Cullen (2010). Estimating Welfare In Insurance Markets Using Variation in Prices. Quarterly Journal of Economics 125(3), 877–921

Main Contribution:

- Estimate the welfare loss caused by adverse selection with reduced form estimation. The key point for this approach is **exogenous price change**.

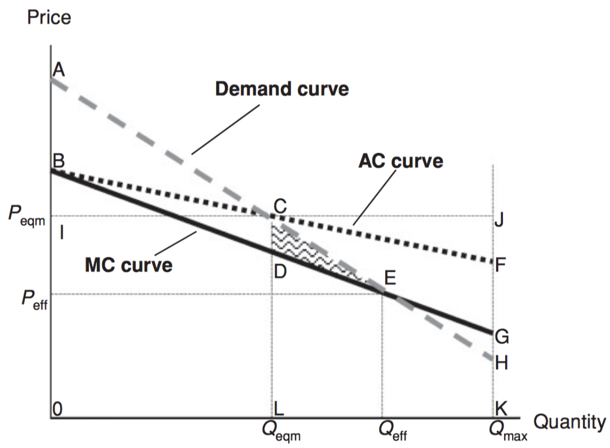


FIGURE I

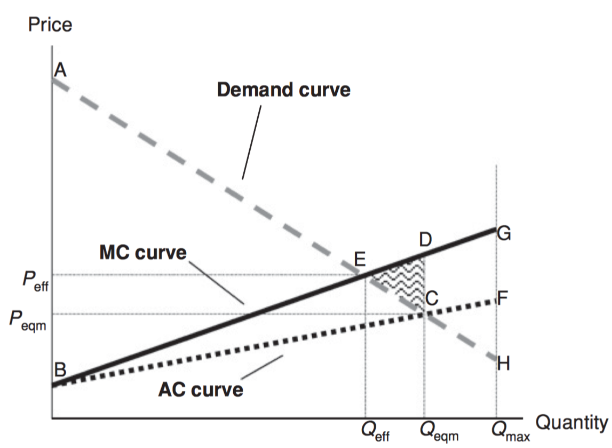


FIGURE II

The adverse selection effect is totally represented by the downward sloping MC/AC curve. If it is advantageous selection, the MC/AC curve will be upward.

Empirical Results:

-

Pros:

- This approach has very low demanding on data. Researchers only need *exogenous price change* to get the MC and demand curve. Such change is easy to obtain in insurance industry.

Cons:

-

Chapter 15

Production Function Estimation

There are several difficulties in estimating production functions which will cause endogeneity problem. To name some:

- Simultaneity. For a production function like $y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \varepsilon_{it}$, we suppose that ω_{it} is obseved by firm but not by econometricians, and ε_{it} is observed neither by firm nor econometricians. This could induce endogeneity when directly using OLS. Suppose ω_{it} has *positive serial correlation*, when ω_{it} is high, firm will choose higher input because it is very possible that ω in next period will be high as well. This induces a positive bias for β_k.
- Exit selection bias. When constructing a balanced panel, then we only observe those firms with high ω_{it}. Why we have to construct a balanced panel?

There are several approaches in solving this problem.

- Classical Approach.
  - Use input price as IV. The problem is input price is generally the same across firms. When we do observe different input prices, which generally indicates market power in input markets. Thus those facing higher ω_i will tend to produce more, inducing their input price to increase, which has a negative bias on β_k.
    - (My comment) When using input price as IV, we actually assume they are independent of individual firm productivities. This may not be the case. If productivity has some correlation across the firms, then it is possible that when productivity is high, the input price in the market will also be high.

- Use fixed effect. We need to assume that a firm faces same ω_{it} across time, which is not very practical.

- Control Function for ω_{it}. This approach first raised by Olley and Pakes (1996) and developed by Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015). We name them OP, LP, ACF here after. The key idea is to use a function of what we know (for example, k_t and i_{t-1} in OP) to proxy ω, the existence of such function is guaranteed by the solutioin property of the firm optimization problem. And the specific form of this function is estimated non-parametrically.
  - OP uses a two step estimation process. First step, use k_t and i_t non-parametrically estimate ω_t, plug this into original production function then use OLS. Notice the unobserved ω_t has been controlled by k_t and i_t, then OLS will give consistent estimator for l_t but not k_t. In second step, we first write $\omega_{it} = E[\omega_{it}|\omega_{it-1}] + \xi_{it}$, notice that ξ_{it} is independent of any information set at t-1. Thus we get the moment condition $E[\xi_{it}k_{it}] = 0$, then we can use method of simulating moments to estimate β_k.
    * The key assumption in first step is, current period investment i_t can be represented by a function of $f(k_t, \omega_t)$. If f is invertible in ω_t (for example, f being monotone in ω), then we can write ω as a function of (i_t, k_t).
  - LP is very similar to OP, just that use intermediate input m_{it} instead of investment i_{it} in proxy ω_{it}. Because OP approach requires the policy function of i_{it} be monotonicity in ω_{it}, which is obviously unsatisfied when $i_{it} = 0$. But data shows that in many observations $i = 0$. Using intermediate input can solve this.
  - ACF criticizing is that β_l cannot be identified in the first step. In LP we assume that $m_{it} = f_t(k_{it}, \omega_{it})$. But we can reasonably argue by the same reason that $l_{it} = h_t(k_{it}, \omega_{it})$. Then since $f(\cdot)$ is invertible, then we can write l_t as some deterministic function of k_{it} and m_{it}. This implies we cannot identify β_l in the first step. The solution of ACF is to identify both β_k and β_l simultaneously at step 2, which is feasible since we now have two moment conditions.

- Gandhi, Navarro, and Rivers (2017) Approach (GNR). GNR criticizes control function approach is nonparametrically not identified in the presence of flexible inputs. They raise a new approach using FOC conditions.

Chapter 16

Multisided Platforms

Schmalensee, D. S. E. and Richard (2013, October). The Antitrust Analysis of Multi-Sided Platform Businesses. manuscript, 1–45

- Price structure matters. Reduce charge by one side and increase charge by another side will affect the transaction.
- Transaction cost is necessarily high for multisided platform.
- Two kind of externalities. Usage externality and membership externality.
- Pricing is compelicated. Per-transaction fee, or membership-fee. This looks similar to conventional IO model.
  - In per-transaction fee case. Price is directly proportional to the elasticity, while in conventional models price is inversely related to elasticity.

$$\frac{p_1 + p_2 - c_1 - c_2}{c_1 + c_2} = \frac{1}{e_1 + e_2}$$

$$\frac{p_1}{e_1} = \frac{p_2}{e_2}$$

    Notice the second equation is unconventional.

  - In membership-fee case.
  - Both models can lead to a highly skewed price structure.
- Competition Structure.
  - Product differentiation

  - Multi-home.
  - Asymetric Competition. A n-sided platform may face competition from 1) one-sided firms; 2) a multisided platform competing in some but not all sides; 3) a multisided platform that has additional sides

---

Random Ideas

- Multisided platform as a commitment device changes the game structure. Like in the Truck-GasStation case raised in the book, the fleet-card company provides commitment power to the card compannny, no renegotiation problem.

Chapter 17

Product Reviews

Berger, J., A. T. Sorensen, and S. J. Rasmussen (2010, September). Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. Marketing Science 29(5), 815–827

Question and Explanations:

- Why negative reviews may have positive effect on selling?
- Negative review may increase awareness of the product.
  - Theory prediction: negative reviews should have more negative effect for high awareness products than for low awareness products.
- Cannot explain: why some established bad movies like *Fu Chun Shan Ju Tu* still sell a lot.
  - May because the different taste between reviewers and mass watchers.
  - As a social topic, just to be fashion.

17.1 Gentzkov-Shapiro-Taddy, 17 WP

Gentzkow, M., J. Shapiro, and M. Taddy (2017, May). Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech

- Methods for analyzing high-dimensional choices decisions
  - also see Gentzkow, M., M. Taddy, and B. T. Kelly (2017, February). Text as Data. pp. 1–53

- Index for measuring polarization

This is a methodology paper introducing how to analyze high-dimensional choices decisions. They show that traditional logit model has severe finite sample bias. They also raise a index for measuring polarization.

[TO BE ADDED]

Chapter 18

Others

Einav, L. and J. Levin (2010, May). Empirical Industrial Organization: A Progress Report. Journal of Economic Perspectives 24(2), 145–162

This survey paper provides a map for current empirical IO research, its range, and its relationship to theory. In addition, the ending part has a critique on the 'reduced' view on IO (e.g. Angrist), worth reading.

18.0.1 Rey and Stiglitz, 1995, RAND

Rey, P. and J. Stiglitz (1995, October). The Role of Exclusive Territories in Producers' Competition. The RAND Journal of Economics 26(3), 431–451

For detailed proof of this paper, see Evernote.

Main result: vertical restrains can be used to reduce interband competition. Because exclusive territories alter the perceived demand curve, making each producer believe he faces a less elastic demand curve, inducing an increase in eqm price and producer's profits even in the abscence of franchise fee. This result is different from traditional Chicago school results, which insist that exclusive terittories will increase efficiency. This difference comes from market structure. Chicago schools investigate in full competition and full monopoly producer cases, while this paper looks at duopoly producer. In full competition and full monopoly case, the competition level has already been *preassumed*, while exclusive territories can reduce the competition level among producers in other cases.

The key for the result is the the following compound demand elastic:

$$\tilde{\varepsilon}(p^e) := m_1(p,p)\varepsilon_1(q,q) + m_2(p,p)\varepsilon_2(q,q)$$

where $q = q_1^r(p,p)$ (the response retail price), $m_1(p,p) = \partial \log q_1^r(p_1,p_2)/\partial \log p_1$ (the own elasticity of retail price to producer price), and $m_2(p,p) = \partial \log q_1^r(p_1,p_2)/\partial \log p_2$ (the cross elasticity of retail price to producer price), and $\varepsilon_1(q_1,q_2) = -\partial \log D^1(q_1,q_2)/q_1$ (the own elasticity of demand to retail price, positive), and $\varepsilon_2(q_1,q_2) = -\partial \log D^1(q_1,q_2)/q_2$ (the cross elasticity of demand to retail price, negative).

In the above equation, it is very reasonable to think $0 < m_1 < 1$ and $0 < m_2$. $m_1 > 0$ because the own elasticity of retail price to producer price is positive. $m_1 < 1$ means that retailer will absorb some increase in producers' price, which will be the case if demand elasticity becomes higher in high retail price. And $0 < m_2$ derives from the two products to be substitutes. Under these two conditions, combined with $\varepsilon_1 > 0$ and $\varepsilon_2 < 0$, then $\tilde{\varepsilon}(p^e) < \varepsilon_1$. Thus, under exclusive territories, the producers' perceived demand curve is less elastic. So the equilibrium price (both retail and producer) is higher even when no franchise fee applies.

Setting:

- two manufacturers produce imperfect substitutes at same marginal cost c
- retailers are perfect competition / or exclusive territory
- the final good demand depends on retail prices and is given by $D^i(q_1,q_2)$
- costs and demand functions are common knowledge, retailers observe all contracts signed by each producer
- producers only observer the quantity bought by retailers; they do not observe the quantities sold by retailers (i.e. fullline forcing is infeasible)
- producers serve many markets at no addtional cost
- consumers have no search cost

Under these settings and information conditions, we can see it as a two stave game. In the first stage, producers simultaneously set wholesale price p_1 and p_2. In the second stage, the retailer observe all wholesale price and decide the retail price simutaneously.

Benchmark Case

We use the following assumptions throughout the paper unless specified otherwise:

1. Let $\pi(p_i, q_1, q_2) := (q_i - p_i)D^i(q_1, q_2)$ denote the retail profit for product i; assume it to be twice differentiable wrt each argument, and is sigle peaked wrt q_i. The reaction function $q_i^a(p_i, q_j)$ is thus continously differentiable and characterized by FOC.

2. Products are substitutes: $\partial D^i/\partial q_i \le 0$ and $\partial D^i/\partial q_j \ge 0$

3. Demand functions are symmetric: $\forall p_1, p_2 \in \mathbb{R}_+, D^1(p_1, p_2) = D^2(p_2, p_1)$

Think of the benckmark case that no vertical restriction so the retailers are perfect competitive, and the producer monopolizes. In this case, the game is just a one step optimal pricing problem. The producer chooses an optimal retail price.

Useful trick:

Throughout this paper, we can write the symmetric eqm conditions in the following form:

$$(p^c - c)/p^c = 1/\varepsilon(p^c, p^c)$$

where p^c is the symmetric eqm price, and ε is some kind of elasticity.

In a symmetric eqm, the FOC of each producer gives $(p^c - c)/p^c = 1/\varepsilon_1(p^c, p^c)$ where $\varepsilon_1 = -\partial \log D^1(q_1, q_2)/\partial \log q_1$, i.e. the self elasticity of demand.

In the simplest benchmark case, the two factories are integrated, leading us to: $(p^c - c)/p^c = 1/E(q^m)$, $E(q) := \varepsilon_1(q, q) + \varepsilon_2(q, q)$, where $\varepsilon_2 = -\partial \log D^1(q_1, q_2)/\partial \log q_2$ (the cross demand elasticity).

In the exclusive territory case, the symmetric eqm satisfies $(p^e - c)/p^e = 1/\tilde{\varepsilon}(p^e)$, where

$$\tilde{\varepsilon}(p^e) := m_1(p, p)\varepsilon_1(q, q) + m_2(p, p)\varepsilon_2(q, q)$$

where $q = q_1^r(p, p)$ (the response retail price), $m_1(p, p) = \partial \log q_1^r(p_1, p_2)/\partial \log p_1$ (the own elasticity of retail price to producer price), and $m_2(p, p) = \partial \log q_1^r(p_1, p_2)/\partial \log p_2$ (the cross elasticity of retail price to producer price).

Part III

Trade and Spatial Economics

Chapter 19

Trade Basic Model

19.1 CES Functions

19.1.1 CES Utility (discrete)

Reference: Rutherford (2002), Lecture Notes on Constant Elasticity Functions

The CES utility function is defined as:

$$U(x,y) = (\alpha x^{\rho} + (1-\alpha) y^{\rho})^{1/\rho}$$

We can calculate the demand:

$$\begin{aligned} x\left(p_x, p_y, M\right) &= \left(\frac{\alpha}{p_x}\right)^{\sigma} \frac{M}{\alpha^{\sigma} p_x^{1-\sigma} + (1-\alpha)^{\sigma} p_y^{1-\sigma}} \\ y\left(p_x, p_y, M\right) &= \left(\frac{1-\alpha}{p_y}\right)^{\sigma} \frac{M}{\alpha^{\sigma} p_x^{1-\sigma} + (1-\alpha)^{\sigma} p_y^{1-\sigma}} \end{aligned}$$

and also the indirect utility:

$$V\left(p_x, p_y, M\right) = M\left(\alpha^{\sigma} p_x^{1-\sigma} + (1-\alpha)^{\sigma} p_y^{1-\sigma}\right)^{\frac{1}{\sigma-1}}$$

V is homogeneous of degree 1 in M. This permits us to form an exact **price index** corresponding to the cost of a unit of utility:

$$\begin{aligned} 1 \equiv V(p_x, p_y, e(p_x, p_y, V=1)) &= e(p_x, p_y, V=1)\left(\alpha^{\sigma} p_x^{1-\sigma} + (1-\alpha)^{\sigma} p_y^{1-\sigma}\right)^{\frac{1}{\sigma-1}} \\ \Rightarrow e(p_x, p_y) \equiv e(p_x, p_y, V=1) &= \left(\alpha^{\sigma} p_x^{1-\sigma} + (1-\alpha)^{\sigma} p_y^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \end{aligned}$$

The indirect utility can be reframed as:

$$V(p_x, p_y, M) = \frac{M}{e(p_x, p_y)}$$

Importantly, wlog, we can think of each consumer demand only one good, and $e(p_x, p_y)$ is the price index for that good.

Elasticity:

CES get its name because the cross elasticity is a constant independent of consumption bundle. To see this we first use a more general form:

$$U(\vec{x}) = \left[\sum_{i=1}^{n} x_i^{\rho}\right]^{\frac{1}{\rho}}$$

By brutal force calculation we get:

$$\frac{\partial x_i / x_i}{\partial p_j / p_j} = \frac{1}{1-\rho}$$

19.1.2 CES Utility (continuous)

CES can be generated to continuous version:

$$\max U(\mathbf{c}) = \left(\int_0^1 b(i)c(i)^{\rho} di\right)^{1/\rho}$$
$$\text{s.t. } \int_0^1 p(i)c(i)di = I$$

Write the Lagranian and we have FOC:

$$b(\iota)\rho c(\iota)^{\rho-1}\frac{1}{\rho}\left(\int_0^1 b(i)c(i)^{\rho} di\right)^{(1-\rho)/\rho} = \lambda p(\iota)$$
$$b(\iota)c(\iota)^{\rho-1}\left(\int_0^1 b(i)c(i)^{\rho} di\right)^{(1-\rho)/\rho} = \lambda p(\iota),\ \forall \iota \in [0,1] \tag{2}$$

Multiply both sides by $c(\iota)$ and take integration:

$$\left(\int_0^1 b(i)c(i)^{\rho} di\right)^{(1-\rho)/\rho}\left(\int_0^1 b(\iota)c(\iota)^{\rho} d\iota\right) = \lambda \int_0^1 p(\iota)c(\iota)d\iota = \lambda I$$
$$\left(\int_0^1 b(i)c(i)^{\rho} di\right)^{1/\rho} = \lambda I$$

Define the LHS as C, the composite consumption, $P \equiv \lambda^{-1}$ and we will get:

$$CP = I$$

we call P the **price index**, which has a natural interpretation to be the price of a unit util. It should be noted that here C is already the *optimal* consumption bundle.

We still need to solve P. Return to (2) and solve $c(\iota)$:

$$c(\iota) = \left[b(\iota)^{-1} \left(\int_0^1 b(i)c(i)^\rho di \right)^{(\rho-1)/\rho} \lambda p(\iota) \right]^{\frac{1}{\rho-1}}$$

$$\begin{aligned} b(\iota)c(\iota)^\rho &= b(\iota) \left[b(\iota)^{-1} \left(\int_0^1 b(i)c(i)^\rho di \right)^{(\rho-1)/\rho} \lambda p(\iota) \right]^{\frac{\rho}{\rho-1}} \\ &= b(\iota)^{\frac{-1}{\rho-1}} \left(\lambda p(\iota)\right)^{\frac{\rho}{\rho-1}} \int_0^1 b(i)c(i)^\rho di \end{aligned}$$

Taking integral in both sides:

$$\begin{aligned} 1 &= \lambda^{\frac{\rho}{\rho-1}} \int_0^1 b(\iota)^{\frac{-1}{\rho-1}} p(\iota)^{\frac{\rho}{\rho-1}} d\iota \\ P \equiv \lambda^{-1} &= \left[\int_0^1 b(\iota)^{\frac{-1}{\rho-1}} p(\iota)^{\frac{\rho}{\rho-1}} d\iota \right]^{\frac{\rho-1}{\rho}} \end{aligned}$$

Properties

Cross Elasticity:

By equation (2), take the log form and take partial derivative, we get:

$$\sigma \equiv \frac{\partial c(x)/c(x)}{\partial p(y)/p(y)} = \frac{1}{\rho - 1}$$

Own Elasticity:

Still from (2), we substitute λ by P^{-1}. Then (2) can be written as:

$$b(\iota)c(\iota)^{\rho-1}C^{1-\rho} = P^{-1}p(\iota)$$

The trick is that a single product c is so small that it will affect neither C nor P. And thus the own elasticity is still $\frac{1}{\rho-1}$.

Limiting Cases:

When $\rho \to 0$, Cobb-Douglas. When $\rho \to 1$, perfect substitution. When $\rho \to \infty$, Leontief. This is from Kim Ruhl's notes, I did not check.

Two Stage Budgeting:

Suppose there is some outside good c_0 in the economy:

$$\max U\left(c_0, \left(\int_0^1 b(i)c(i)^\rho di\right)^{1/\rho}\right)$$
$$\text{s.t. } p_0c_0 + \int_0^1 p(i)c(i)di = I$$

One thing important is that the differentiated goods can be aggregated into one commodity C. Then the above problem can be summarized as a two stage problem.

$$\max U\left(c_0, C\right)$$
$$\text{s.t. } p_0c_0 + PC = I$$

We first solve C and PC. Then everything comes similar.

19.2 Ricardian Model

This representation is based on DFS 1977.

Settings:

- Two countries: Home and Foreign
- One factor of production
  - This is important in Ricardian model, because the order of z below relies on unique factor of production. Otherwise we won't have a meaningful order.
  - Denote L, L^* as the labor endowments in two countries; and w, w^* be the wages respectively. All terms are real.
- A continuum of goods indexed by $z \in [0, 1]$
  - A continuum of z is easier to deal with than discrete z.
- Technology: $a(z)$ and $a^*(z)$ are the labor requirements for produing unit good.

– Order goods s.t. $A(z) \equiv \frac{a^*(z)}{a(z)}$ is decreasing. So Home has comparative advantage in low z.

Equilibrium:

- Suppose there is no trade cost. Let $p(z)$ denote the (home) price of good z under free trade.
- Since the competition is perfect, and there is no trade cost, the product will be produced in the country with lower labor cost:

$$p(z) - wa(z) \leq 0, \text{w . equality if } z \text{ is produced at Home (1)}$$
$$p(z) - w^*a^*(z) \leq 0, \text{ w. equality if } z \text{ is produced Abroad (2)}$$

- A direct proposition is: there exists $\tilde{z}$ s.t. Home produces all goods $z < \tilde{z}$ and vice versa for Foreign. Another result is: $A(\tilde{z}) = \frac{w}{w^*} \equiv \omega$. Note that ω is endogenous and can be calculated in equilibrium.
- To calculate ω, we further assume consumers have **identical Cobb Douglas** preference around the world. Then we denote $b(z)$ the share of expenditure on good z: (which actually should be understand as density of share):

$$b(z) = \frac{p(z)c(z)}{wL} = \frac{p(z)c^*(z)}{w^*L^*}$$

by def, shares should satisfy:

$$\int_0^1 b(z)dz = 1$$

Let $\theta(\widetilde{z}) \equiv \int_0^{\widetilde{z}} b(z)dz$ denote the fraction of income spent on goods produced at Home. Trade balance requires Home Exports = Home Imports:

$$\theta(\widetilde{z})w^*L^* = [1 - \theta(\widetilde{z})]wL$$

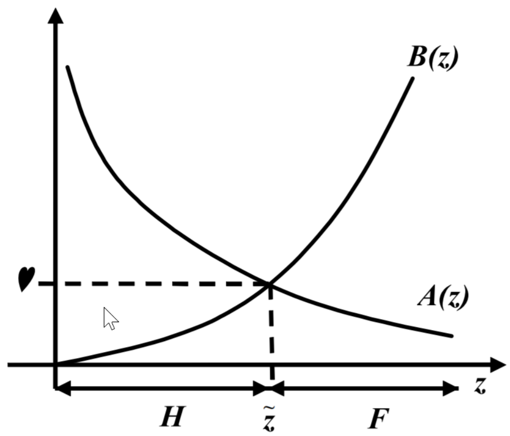
then we have relative wage to be:

$$\omega = \frac{\theta(\widetilde{z})}{1 - \theta(\widetilde{z})} \left(\frac{L^*}{L}\right) \equiv B(\widetilde{z})$$

Notice this is an GE results. An technical advancement will increase $\tilde{z}$, which must be compensated by an increase in ω in equilibrium.

Comparative Statics:

- Below figure plots the equilibrium.$A(z)$ is downward by construction, $B(z)$ is upward by above equation.



- Suppose L^*/L increases, how it will change equilibrium price and welfare?
  - For notation ease, set $w = 1$ wlog since it only serves as numerair.
  - First notice that ω increases and $\tilde{z}$ decreases in equilibrium.
  - For goods produced at Home, no change in $p(z)$ since it only depends on Home technology.
  - For goods whose production remains at Abroad:

$$\omega \nearrow \Rightarrow w^* \searrow \Rightarrow p(z) = w^* a^*(z) \searrow$$

  - For goods moved to be produced Abroad, then by profit optimization:

$$w^* a^*(z) \leq a(z) \Rightarrow p(z) \searrow$$

  - Thus Home gains.
- KZ: The Abroad welfare decreases because of population increase. Why does this happen? Because Abroad now produces goods that they have less comparative advantages. They produce these goods because they have low wage, due to the competition effect of large population.

19.3 Eaton and Kortum, 02 EMCA

Setting:

- Efficiency: country i's efficiency in product j is $z_i(j)$
- Input price in country i is c_i, which can be further decomposed into labor cost and intermediate inputs cost

- Geographic barriers between country i and country n is d_{ni}. Is d_{ni} symmetric?

- Production cost is:
$$p_{ni}(j) = \left(\frac{c_i}{z_i(j)}\right) d_{ni}$$
which is also the purchase cost by perfect competition assumption. However, the buyer will buy from the cheapest seller, and we assume the search is costless:
$$p_n(j) = \min \{p_{ni}(j); i = 1, \ldots, N\}$$

- Utility: buyers have the love of variety, and we assume a CES utility form:
$$U = \left[\int_0^1 Q(j)^{(\sigma-1)/\sigma} dj\right]^{\sigma/(\sigma-1)}$$
where the elasticity of substitution is $\sigma > 0$.

Technology:

- The technology $z_i(j)$ is the realization of r.v. Z_i, drawn independently for each product j. Z_i follows a Type II extreme value distribution:
$$F_i(z) = e^{-T_i z^{-\theta}}$$
T_i governs the location of distribution, and θ common for all countries governs the variation *within* the distribution. A bigger θ represents less variability. This distribution is the key to the closed form solution.

Prices:

- The induced price distribution is:
$$\begin{aligned} G_{ni}(p) &= \Pr\left(P_{ni}(j) \leq p\right) \\ &= \Pr(\frac{c_i}{z_i(j)} d_{ni} \leq p) \\ &= 1 - F_i(\frac{c_i d_{ni}}{p}) \\ &= 1 - e^{-T_i(c_i d_{ni})^{-\theta} p^{\theta}} \end{aligned}$$

The purchasing price distribution is thus: how to derive?

$$
\begin{aligned}
G_n(p) &= 1 - \prod_{i=1}^{N} [1 - G_{ni}(p)] \\
&= 1 - e^{-\Phi_n p^\theta}
\end{aligned}
$$

where Φ_n is:

$$\Phi_n = \sum_{i=1}^{N} T_i \left(c_i d_{ni}\right)^{-\theta}$$

- The probability that country i provides a good at the lowest price in country n is:

$$
\begin{aligned}
\pi_{ni} &= \Pr\left[P_{ni}(j) \leq \min\left\{P_{ns}(j); s \neq i\right\}\right] \\
&= \int_0^\infty \prod_{s \neq i}^{\infty} [1 - G_{ns}(p)]\, dG_{ni}(p) \\
&= \frac{T_i \left(c_i d_{ni}\right)^{-\theta}}{\Phi_n}
\end{aligned}
$$

the last step?

- The price of a good that country n *actually buys* from country i also has the distribution $G_n(p)$. This means that the price distribution *conditional* on export country is the same as overall price distribution faced by importer country n. This result can be derived by showing that:

$$G_n(p) = \frac{1}{\pi_{ni}} \int_0^p \prod_{s \neq i} [1 - G_{ns}(q)]\, dG_{ni}(q)$$

- The price index for CES function is then (why?)

$$p_n = \gamma \Phi_n^{-1/\theta} \tag{9}$$

where

$$\gamma = \left[\Gamma\left(\frac{\theta + 1 - \sigma}{\theta}\right)\right]^{1/(1-\sigma)}$$

Γ is the gamma function.

Trade Flows and Gravity

- Since the price distribution conditional on source is the same as overall price distribution, a corollary is that the average expenditure per good is identical across sources. Thus, the fraction of goods that country n buys from country i, π_{ni} is the fraction of expenditures:

$$\frac{X_{ni}}{X_n} = \pi_{ni} = \frac{T_i\left(c_i d_{ni}\right)^{-\theta}}{\Phi_n} = \frac{T_i\left(c_i d_{ni}\right)^{-\theta}}{\sum_{k=1}^{N} T_k\left(c_k d_{nk}\right)^{-\theta}} \tag{10}$$

The exporter's total sales Q_i are:

$$Q_i = \sum_{m=1}^{N} X_{mi} = T_i c_i^{-\theta} \sum_{m=1}^{N} \frac{d_{mi}^{-\theta} X_m}{\Phi_m}$$

Then we can cancel out $T_i c_i^{-\theta}$:

$$T_i c_i^{-\theta} = \frac{X_{ni}\Phi_n d_{ni}^{\theta}}{X_n} = \frac{X_{ni}}{X_n}\left(\frac{p_n}{\gamma d_{ni}}\right)^{-\theta}$$

the second equation is from derived CES price index equation. Substitute $T_i c_i^{-\theta}$ into Q_i and we get:

$$X_{ni} = \frac{\left(\frac{d_{ni}}{p_n}\right)^{-\theta} X_n}{\sum_{m=1}^{N}\left(\frac{d_{mi}}{p_m}\right)^{-\theta} X_m} Q_i \tag{11}$$

Why we are interested in this equation? Because it resembles the standard gravity equation.

Trade, Geography and Price: The Connection:

- Notice that if we divide (10) by X_{ii}/X_i, we get:

$$\frac{X_{ni}/X_n}{X_{ii}/X_i} = \frac{\Phi_i}{\Phi_n} d_{ni}^{-\theta} = \left(\frac{p_i d_{ni}}{p_n}\right)^{-\theta}$$

the LHS is called *normalized import share*. why we want to normalize it?

Input Costs:

Production:

- The cost of an input bundle in country i is:

$$c_i = w_i^{\beta} p_i^{1-\beta}$$

Part IV

Micro Theory

Chapter 20

Game Theory

20.1 Important Tools

Theorem 20.1.1 (Revelation Principle, Myerson, 1981). *Suppose that ψ was a Bayes Nash equilibrium of the indirect mechanism Γ. Then there exists a direct mechanism that is payoff-equivalent and where truthful revelation is an equilibrium.*

Proof: [to be added] $\square$

Comment:

- Revelation Principle is defined over some specific solution concept. Here it is Bayes Nash Eqm.
- Revelation principle is not always satisfied. The stronger solution concept we use, the more likely it is not satisfied.

20.1.1 Solution Concept

A very good reference for different solution concept is van Damme (1987).

- Nash eqm
- Subgame Perfect eqm
- Perfect eqm

- ε-Perfect eqm
- Proper eqm
- Trembling-Hand eqm
- Bayesian Nash eqm
- (Weak) Perfect Bayesian eqm
- Sequential eqm
- Forward looking eqm
- Self-Confirming eqm

Definition 20.1.1 (Self-confirming Equilibrium, Fudenberg and Levine (1993)). Profile σ is a *self-confirming equilibrium* if for each player i, $s_i \in \text{support}(\sigma_i)$ there exists belief μ_i such that:

1. s_i maximizes $u_i(\dot{;}\mu_i)$, and
2. $\mu_i[\{\pi_{-i}|\pi_j(h_j) = \hat{\pi}_j(h_j|\sigma_j)\}] = 1$, for all $j \neq i$ and $h_j \in \bar{H}(s_i, \sigma_-i)$, where $\bar{H}$ is the set of info sets that are reached with positive probability

Comment:

- This definition mainly requires that the belief μ_i is consistent only at information sets that are reached with positive probability when player i plays s_i.
- This is most obvious when contrast with Nash Eqm, which is almost the same as this definition, except for the consistency requirement is for all $h_j \in H_{-i}$. The consistency is required for all information sets.
- Self-confirming is a weaker solution concept than Nash Equilibrium. It uses the idea that players can only learn through what they observe. For those outcomes that they cannot observe, there is no particular reason for their beliefs to coincide. Players can never learn something they lack information on.

Example 20.1.1.

20.2 Decision Theory

20.2.1 Morris, 95 Econ&Phil

Morris, S. (1995, October). The Common Prior Assumption in Economic Theory. Economics and Philosophy 11(2), 227–253

This assumption reviews the common prior assumption in economics, its common justifications, and its impact on game theory.

VEEY WORTH READING!! NOTES TO BE ADDED.

20.3 Mechanism Design

Important tools and topics in mechanism design:

- Revelation Principle
- Single dimension mechanism design
  - Bayesian mechanism design
  - Dominant strategy mechanism design
  - Revenue equivalence
  - Payoff equivalence
- Multidimensional mechanism design
- Interdependent types mechanism design

20.3.1 Revelation Principle

There are many things assumed in the proof of revelation principle. Take p.10 in Borger's book as example.

- No verification.
- Utility is transferable.

For an example where all three assumptions are violated, see Ben-Porath, Dekel, and Lipman (2014).

Information side of Revelation Principle:

In setting we prove revelation principle, we assume every thing is common knowledge except for the specific type. These include the type space, the action space of each type, the utility form of each type, and so on. We also assume that in the direct mechanism, the action space is the *whole* type space for *every* agents. [Don't understand this. Since in the proof we only require that one can report truthfully.]

20.4 Learning Theory

20.4.1 Important tools in learning

Definition 20.4.1 (Monotone Likelihood Ratio Property). For all $l < S$,

$$\frac{p_{q,l}}{p_{q+1,l}} \geq \frac{p_{q,l+1}}{p_{q+1,l+1}} \ for \ all \ q < R$$

with strictly inequality for at least one q.

This is first raised by Milgrom (1981), which ensures that the conditional expectation of each individual increases in his signal realization.

Definition 20.4.2 (Martingale). A *martingale* is a pair $(X_n, \mathscr{F}_N)$, where $\{\mathscr{F}_n\}$ is a filtration and X_n an integrable (i.e. $E|X| < \infty$) stochastic process adapted to this filtration, such that

$$E[X_{n+1}|\mathscr{F}_n] = X_n, \ a.s. \ \forall n$$

Comment:

- Notice that Martingale process is much more restricted than Markov process, not only in the realization of expectation value. Markov process requires that

  $$Pr(X_{t+1} \in A)|X_t = x_0, X_{t-1} = x_{t-1}, ..., X_{t-k} = x_{t-k} = Pr(X_{t+1} \in A)|X_t = x_0$$

  for all $A \in \mathscr{X}$, all $x_0, x_1, ..., x_k \in \mathbb{X}$, and all $0 \leq k \leq t$.

  - This actually restricts all stochastic properties, including expectations, variations and others.
  - Also, Martingale process require conditional on $\mathscr{F}_t$, but Markov process only requires x_t.

Theorem 20.4.1 (Martingale Convergence Theorem).

20.4.2 Overview of literature

The equilibrium concept used in literature is Bayesian Nash Eqm. Thus we put no restriction on off-eqm path beliefs. Nageeb said something about this, but I don't understand why this is important.

Bikhchandani et al. (1992) (BHW here after) is one of the original papers for learning in games. The main point is to explain the easily formulated but fragile information cascade when only actions are observed. The idea is, if only actions are observed, then after some threshold, private information will not affect action. And this private will NOT join the information pool, thus all following individual's private information will not contribute to the action. But since very little information is needed to form a cascade, this cascade can also be easily reversed when some shocks happen.

Smith and Sørensen (2000) is a generalization of BHW paper. The generalization lies in these aspects:

- Heterogeneity in quality of information (possibly unbounded private information)
- Heterogeneity of preference (noise by crazy types)
- Information cascade is NOT a generic feature; but 'herding' is.
- Necessary and sufficient condition for *complete learning*.

This paper also provides a general approach on how to handle learning behaviour.

20.4.3 Bikhchandani-Hirshleifer-Welch, 92 JPE

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. Journal of Political Economy 100(5), 992–1026

BHW is the seminal paper of social learning, which explains imitating in social phenomena without any ad hoc assumptions. By ad hoc assumptions, I means following:

- sanctions on deviants
- positive payoff externalities
- conformity preference

This assumptions all have the unsatisfying property that puts the cart before the horse.

TO BE ADDED

20.4.4 Kremer-Mansour-Perry, 14 JPE

Kremer, I., Y. Mansour, and M. Perry (2014). Implementing the "Wisdom of the Crowd". Journal of Political Economy 122(5), 988–1012

Different from previous learning papers, KMP focuses on the optimal information disclosure of planner. They find that it is not always optimal for planner to disclose full information, because it will discourage agents' incentive to explore and exploration has positive externalities.

Example: Online review website like TripAdvisor, Yelp; Google Map

Illustrative example:

Say R_1 is uniformly distributed in $[-1, 5]$, and R_2 is uniformly distributed in $[-5, -5]$. Then agent-1 for sure will choose a_1 since its ex-ante expeted payoff is higher. Planner cannot force agent-1 to choose a_2 because of IC condition.

If realized R_1 is lower than 0, then for sure agent-2 will choose a_2. But planner can do better by recommending agnet-2 to play a_2 whenever $R_1 \leq 1$. Agent-2 will follow this suggestion because the $E[R_1]$ is 0 conditional on planner not disclosing.

If unfortunately the realized R_1 is higher than 1, then there is no way to make agent-2 to choose a_2, but it is still possible to ask agent-3 to explore if R_1 is not too high. Because when player-3 is recommended to play a_2, she cannot distinguish between: 1) planner knows both realization and $R_2 > R_1$; or 2) planner wants her to explore. The second situation is in the same shoes as agent-2, but because of the possibility of the first situation, agent-3 will admit a higher upper threshold for exploring.
□

Main Results: (The proof of this paper is not difficult, and worth learning)

- The optimal policy is threshold policy. For each individual, there is an interval of realization r_1 such that he is the first agent to explore.
- There is an upper bound of exploration time, which is independent of agents number T and realization of R_1, R_2. This indicates result will converge to first best when T is large enough.

Comments:

- Generalize action to n will not affect the structure of optimal policy. While continuum of action will ask for new setting. See referee report for more.

- We can also think the realization of R_1 and R_2 is **correlated**. The result of this paper will not hold in this case. Think of the extreme case that correlation between R_1 and R_2 is -1. Then for any $R_1 > 0$, the planner will never explore a_2. For any $R_1 < 0$, the planner will recommend a_2 for all agents except for first one, and they will accept this recommendation. The correlation problem also exists when the relization of R_i is not a value but a distribution, in the sense of correlation in mean value. In another direction, if the realization is positively correlated, then the exploration cost will be lower. And optimal policy should have lower upper threshold for each agent.

  I think this extension important, because we have correlations in many real world applications. Take TripAdvisor as an example. If we already have reviews on one Marriott hotel in Paris, then it's not very useful to explore another Marriott at Toulouse! Intuitively, planner should utilize such correlation, and put more weight on exploration of independent realizations. The optimal policy could be interesting.

- I also think of how this can be combined with optimal persuasion. If planner and agents have discrepancy in interest, then persuasion is needed.

20.5 Hard Information

Unraveling:

Unraveling is first discussed by Grossman (1981) and Milgrom (1981). They show that if investor knows the type of firm's information, such information is perfectly verifiable at no cost, and the disclosure of such information is at no cost. Then exist a sequential equilibrium, where the firm will disclose all information, good or bad. The intuition is, since investors know the type (distribution) of information firm has, the nondisclosure indicates a worst information. If the investor is so suspicious, firm will disclose those not so good information.

Dye (1985) finds that such unraveling may not always happen. There are many possible reasons, basically they are violations of above three assumptions.

Quigley and Walther (2017) discusses unraveling in existence of outside information. They show that such information beyond control of insiders will *reverse* unraveling result. The highest quality insiders remain quiet, and only the mediocre insiders disclose information. Notice that sending message is costly in this model. The intuition for this result is, when type is high enough, outside information is enough to persuade the outsider. My guess is, this holds because the outsider decision is not fully divisible.

Mechanism Design with Hard Information:

Glazer and Rubinstein (2004) discusses the optimal decision rule when listener can check only one value. [Not yet read]

Green and Laffont (1986) discusses the direct mechanism under partially verifiable information. They show that the *Nested Range Condition* is necessary and sufficient for all socially-implementable mechanisms to be truthfully-implementable mechanisms. The condition requires that if A can pretend to be B, B can pretend to be C, then A can pretend to be C. This condition is used in many other papers like Hart et al. (2017), and monotonicity condition in Grossman (1981) is also included. But notice that their paper does not consider a explicit checking process, so the result cannot apply to Ben-Porath et al. (2014), where a perfect but costly checking technique exists.

20.5.1 Evidence Game

[Below is based on the talk by Xiao] The paper by Hart et al. (2017) provides a general framework for evidence games. It shows that under conditions:

- A1. $\theta \in m(\theta) \subseteq \Theta$. It basically restricts the message space to be the type space, and every can truthfully report his own type.
- A2. If $\theta' \in m(\theta'')$ and $\theta \in m(\theta')$, then $\theta \in m(\theta'')$. It means the masquerade is transitive, which is quite restrictive condition.

With above two assumptions, and *truth-leaning* refinement of equilibriums, they show that giving receiver commitment power will make him no strict better off.

Many more basic models like Dye (1985) and Milgrom (1981) can be encompassed by this framework. [Worth trying to put them in.]

20.5.2 Beyer-Cohen-Lys-Walther, 10 JAcctE

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2010, December). The financial reporting environment Review of the recent literature. Journal of Accounting and Economics 50(2-3), 296–343

This paper surveys literature of **voluntary information disclosure**, which is highly correlated with *hard information* literature we discuss in reading group. I mainly read the section 3 (voluntary disclosure) of this paper.

We have unraveling result in Milgrom (1981) and Grossman (1981). Unraveling means the agent will disclose all their private information. This result relies on the following conditions:

1. disclosures are costless;
2. investors know that firms DO have private info;
3. all investors interpret the firm's disclosure in the same way, and firms know how investors will interpret that disclosure;
4. agents want to maximize their firm's share prices (no moral hazard problem);
5. agents can credibly disclose their private info;
6. firms cannot commit ex-ante to a specific disclosure policy.

Violation of any conditions will result in partial nondisclosure of private information.

1. *Costly disclosure.* If disclosure is costly to a firm, then it will choose not to disclose even when signal is not that bad.
2. *Probabilistic information endowment.* If firms get private information in probability, then they have incentive to not report some bad results. Because investors cannot distinguish them from those without private information.
3. *Uncertain investor response.*
4. *Uncertain disclosure incentives.* In this case, there will be moral hazard problem.
5. *Non-verifiable disclosure.* There are many important models in this category. For example, Cheap Talk model and Costly State Verification models. In cheap talk, exists "babbling eqm" since signal can't be verified. In Costly State Verification, firms have some rent since it's not wise to investors to check at every case.
6. *Ex-ante commitment.*

20.5.3 Milgrom, 81 Bell

Milgrom, P. R. (1981, October). Good News and Bad News: Representation Theorems and Applications. The Bell Journal of Economics 12(2), 380–391

Milgrom (1981) introduces a way to measure how good a message is. The difficulty here is that we are comparing two conditional distributions, not two values, thus something similar to FOSD is needed.

Definition 20.5.1. A signal is *more favorable than* another signal if the posterior belief $G(\cdot|x)$ FOSD $G(\cdot|y)$.

Proposition 20.5.1. x is more faborable than y iff for every $\theta_1 > \theta_2$,

$$\frac{f(x|\theta_1)}{f(x|\theta_2)} > \frac{f(y|\theta_1)}{f(y|\theta_2)} \tag{20.1}$$

Intuitively, x is more favorable iff x is more likely to appear when the state is high.

Definition 20.5.2 (Strict Monotone Likelihood Ratio Property). The family of densities $\{f(\cdot|\theta)\}$ have the monotone likelihood ratio property if for every $x > y$ ad $\theta_1 > \theta_2$, (20.1) is satisfied.

Densities with MLRP has the good property that any two signals are comparable.

20.6 Bayesian Persuasion

The question that *Information Design* wants to answer is corresponding to that of Mechanism Design. In mechanism design, the designer can commit to the *rule* of a game; in information design, the designer (sender) commits to a information structure.

20.6.1 Kamenica and Gentzkow, 11 AER

Kamenica, E. and M. Gentzkow (2011, October). Bayesian Persuasion. American Economic Review 101(6), 2590–2615

Example 20.6.1. Suppose there is one prosecutor (Sender) and one judge (Receiver). Both parties hold the prior that the suspect is guilty with $\rho = 0.3$. The prosecutor decides whether to investigate the suspect. After that the judge will see investigation report, which cannot be manipulated by the prosecutor, and then make the decision of acquit or convict. The prosecutor knows the world state (innocent or guilty), but he only wants to increase the probability of convict. The judge wants to matching state. What is the best thing the prosecutor can do?

If the prosecutor does nothing, then the judge will action according to his prior, and *acquit* the suspect with probability 1 since $\rho < 0.5$.

But the prosecutor can do much better than this. Suppose the prosecutor commits to following investigation:

$$\pi(i|innocent) = 4/7, \;\; \pi(g|innocent) = 3/7$$

$$\pi(i|guilty) = 0,\ \pi(g|guilty) = 1$$

Thus when the judge see signal i, he does Bayesian inference and gets: $Pr(guilty|i) = 0$ and $Pr(guilty|g) = \frac{1*0.3}{1*0.3+3/7*0.7} = 0.5$. If the equilibrium is in favor of the prosecutor, the judge will choose *convict* with probability 1 when seeing signal g. [1] Then the total probability of *convict* becomes 0.6.

Comment to Example:

- The sender actually commits to the *precision* of investigation. The point is that he cannot change the result *ex post*.
- The judge will behave the same whether he believes the guilty probability is 0 or 0.49. So the information revelation scheme will make the belief of guilty under signal i to be as low as possible, and make the belief under g slightly over 0.5.

20.6.2 Gentzkow and Kamenica, 17 RES

Gentzkow, M. and E. Kamenica (2017, January). Competition in Persuasion. The Review of Economic Studies 84(1), 300–322

Highlight of the Paper:

- This paper shows that the competition does not increase the amount of information being revealed in general. A key condition for competition to be beneficial is **Blackwell-connectedness**.
- There are other papers answering this question. For example, the cheap talk games and the hard evidence game. But this paper is different because those papers take senders' information as given, and focus on the strategic communications subject to incentive compatibility. This paper abstracts from the incentive problem by giving senders commitment and takes the senders' information structure as endogenous.

20.7 Repeated Game

This section highly depends on Nageeb's notes and Mailath-Samuelson book.

[1] This specific tie breaking rule is not important since we can always add ε to 0.5.

20.7.1 Imperfect Monitoring

- Nageeb's notes

Motivation for imperfect monitoring:

- In practice, monitoring is often imperfect (Green and Porter (1984))
- NOT UNDERSTAND. Helps to understand punishments at off-path histories

With imperfect monitoring, 'forgiveness' will often make the equilibrium better while keeping IC satisfied.

The difficulty in describing imperfect monitoring game with automata is, we can only observe ex-post history, how can we use automata graph to describe this ex-post history? The answer is we can think the equilibrium (and strategy) ex-ante. Because everything is linear, so it actually does not really matter what ex post realization is.

20.7.2 Miscellanea

Hadfield, G. K. and B. R. Weingast (2012, November). What Is Law? A Coordination Model of the Characteristics of Legal Order. Journal of Legal Analysis 4(2), 471–514

This paper asks an interesting question: what is the difference between law and social norm? And they argue that the key feature of law of 'central judgment', while the punishment can be decentralized.

Such a model wants to answer: what's the incentive for individuals to conduct punishment despite that such punishment is costly? Their model is weird.

KZ: Intuitively this is like private monitoring.

20.8 Financial Economics

Bid-Ask Spread:

O'hara (1998) Chapter-3 provides two explanations on existence of bid-ask spread, even with market makers being risk neutral and competitive. Both model based on asymmetrically informed investors.

Copeland and Galai (1983) uses a one-stage game with asymmetric informed investors. The informed investors have full information, while the uninformed investors have no information except for common prior. The market maker will lose money when transact with informed with informed investors, and earn money when transacting with uninformed ones. The bid ask spread is to make market makers balance off.

Several things are unspecified in their model. First, the motivation of uninformed participants are unspecified. Why they will enter the market knowing that they will lose money in expectation? Secondly, the information conveyed by trade action is not specified. A sell order may represent a bad news of informed investor and vice versa.

Glosten and Milgrom (1985) addresses the second concern by studying a sequential trade structure. Market makers will learn in Bayesian to update his belief about the world state. And the buy/sell order conveys different information about the world state since the existence informed investors. So when the investor wants to sell a stock, market maker will have a more gloomy prediction about world state, and thus decrease the price; vice versa.

Glosten-Milgrom model explains the bid-ask spread in abscence of any exogenous transaction or inventory cost. In addition, they show that the price is a Martingale (for market maker). Their model also suggests that market may collapse if there are too many informed investors, since the spread in positively correlated with portion of informed investors.

KZ: But in their model, investors will not use the information of previous transactions, which is unsatisfactory. We can think in social learning approach. If that is the case, then transaction conveys no additional information if investors are in a herd.

20.9 Seminar Notes

George Mailath (WP 17): "The Wisdom of The Confused Crowd: Model-Based Inference"

Main idea:

- If players are non-Bayesian, how do they update? How they exchange information in the information mkt?
- Agents do 'model-based' inference. Basically this is a two step structure.
- Agents have a model which addresses on some dimension of the information, and when they receive an

information, they will only first focuses on the dimensions that their model addresses.

- Then they exchange the posterior belief with other agents. Since other agents have different model addressing different aspects, their posterior belief reveals some information of other dimension signal.
- Since the signals in different dimensions are correlated, one agent may use other's posterior belief to infer partial information of signals in the dimensions that he is interested in.
- Thus, although they have different models addressing different dimensions, they can still exchange in the info market.

Simon Board, Moritz Meyer-ter-Vehn (WP 17): An Inspection Model of Learning in Networks

Main idea:

- Note of Talk
- Two concerns in network learning
  - I get awake at a constant rate, and may want to try
  - I see my neighbor trying, and want to try
- Learning in small network is very difficult to numerically computing.
  - Self-reinforcement structure
  - Triangle correlation
  - $\Rightarrow$ can characterize, but cannot compute
- In large random network, such concern disappears
  - Because in large random network, the prob of each two points are connected is the same. And with large enough network, the prob of a small local structure is second order small.
  - Can write the graph as a tree in this case
- (?) Learning rate increases exponentially with the number of neighbors.

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2017). Revealed price preference: Theory and stochastic testing

Important paper, worth reading.

This paper raises a new representation theorem over preference and utility. Traditionally we use preferences over consumption bundles x, and Afriat theorem guarantees a well-behaving utility representation if GARP is satisfied for preference. This paper raises a representation theorem for preference over price bundles.

More specifically, GARP says if we have $p^{t'} \cdot x^t < p^{t'} \cdot x^{t'}$, then we conclude $x^{t'}$ is revealed preferred to x^t, because x^t is affordable when the consumer choose $x^{t'}$. To test this approach, econometricians have to observe the consumption bundles of *all* goods, which is infeasible in many cases.

This paper raises another way to view data.

- If $p^{t'} \cdot x^t < p^t \cdot x^t$, then we say $p^{t'}$ is revealed preferred to p^t. Because the consumer has money left for other goods when price is $p^{t'}$. In this way, we can construct the preference over prices.
- This paper raises GAAP corresponding to GARP, and augmented utility function (seems integrating expenditure directly in the utility) w.r.t. GAAP.
- This representation can be empirically tested without observing all goods being purchased. And solves the endogenous expenditure level problem, which McFadden-Richter estimation approach assumes.
- Ron points out that this representation model cannot account for budget change.

Yuval Salant (18 WP): Statistical Inference in Games

- Idea: the agents in the game only use a sample of all agents, and based on the decision of the sample, the agent forms an inference. Then makes his own decision based on this inference.
- The trade-off the agent is making is: enlarging sampling size is costly, but a larger sampling size will induce more precise distribution, which is beneficial since agents are risk averse.
- They proposes a *Generalized Sampling Equilibrium* based on above idea. And they show that this equilibria is always smaller than Nash.
- The difficulty in defining this equilibrium is that everyone draws a different sample, and how to define equilibrium in this case? The trick here is they use a continuum of agents, so that even though everyone draws differently, but as a population they draw a same "distribution". The difference among agents are just changing labels.
- Applications:
  - In defining equilibria with search behavior.

  - Wei stated that this concept can be used widely in macro. For example, the shock in TFP will induce an endogenous fluctuation in sampling size; so the total variation is enlarged.

Cooperation Game:

Nageeb mentions in class that one reason for cooperation game to be unpopular is the difficulty in treating informations. How to treat to coalition when someone has information while others don't?

Fainmesser, I. P. and A. Galeotti (2018). The Market for Influence. SSRN Electronic Journal, 1–40

This paper discusses how how

Chapter 21

Matching Theory

21.1 Gale and Shapley, 1962, Ame.Math.Mon.

Gale, D. and L. S. Shapley (1962, January). College Admissions and the Stability of Marriage. The American Mathematical Monthly 69(1), 9–15

This paper introduces Deferred Acceptance (DA) algorithms in matching, and shows that it is stable and optimal.

Definition 21.1.1. An assignment is called **unstable** if there are two applicants α and β who are assigned to universities A and B, although β prefers A to B and A prefers β to α.

Definition 21.1.2. A stable assignment is called **optimal** if *every* applicant is at least as well off under it as under any other stable assignments.

Comment: it looks at first that optimal assignments may not always exist. But as we will show constructively, it does.

Definition 21.1.3. Deferred Acceptance algorithm works as following. First let each boy propose to his favorite girl. Each girl who receives more than one proposal rejects all but her favorite boy. Yet, she doesn't accept him, but keeps him in a string to allow for the better may come later. Then in the second round, each rejected boy propose their second favorite girl, and each girl picks her favorite among the new proposers and the one in the string, and put him in her string. This process continues, until each girl has one boy in the string (this will happen after finite rounds), then each girl picks the boy in her string.

We will show the assignment by above algorithm is stable and optimal.

Part V

Macro

Chapter 22

Bubbles

22.1 Milgrom and Stokey, 82 JET

Milgrom, P. and N. Stokey (1982). Information, trade and common knowledge. Journal of Economic Theory 26(1), 17–27

Theorem 22.1.1 (No Trade Thm). *Suppose that all 1) traders are risk averse; 2) initial allocation $e = (e_1, ..., e_n)$ is Pareto optimal relative to θ-trades; 3) agents' prior beliefs are concordant; 4) each trader observes private information conveyed by the partition $\hat{P}_i$. If it is common knowledge at some world state ω that t is a feasible θ-trade and each trader weakly prefers t to the zero trade, then every agent is indifferent between the two. If all agents are strictly risk averse, then t is zero trade.*

Comment:

- This being called *no trade theorem* because private information will not lead to trade, which is on the contrary to what static models predicts.
- Intuitively, if some other wants to trade with me, then he must possess some private information that in favor of this trade. But if we have same concordant prior belief, such information must *in against with* my trading with him. So I will not accept any trade being offered to me. By the same argument, others will not accept any offer being offered to them.
- World state space is: $\Omega = \Theta \times X$. Θ is set of payoff-related events, which will affect endowment and utility (type), X will not affect endowment and taste directly. A trade is a function from Ω to R^{ln}, it is θ-trade if it only depends on θ.

- Beliefs are *concordant* if:

$$p_1(x|\theta) = ... = p_n(x|\theta),\ \forall x, \theta$$

22.2 Geanakoplos and Polemarchakis, 82 JET

Geanakoplos, J. D. and H. M. Polemarchakis (1982). We can't disagree forever. Journal of Economic Theory 28(1), 192–200

Main result:

- **Theorem:** With common priors, and if the information partitions of two agents are finite, then they can converge to a common equilibrium posterior by communicating *only posterior*. And the communication steps are finite.
- This result is more general than Aumann's, because it does not require the event A to be posterior common knowledge. Actually it does not put any restriction on event A.
- Such converging posterior may NOT be the same as the one they will reach if they direct communicate.
- Such convergence may NOT happen in *each* step. It is possible that the posterior of both agents do not change until the last step. E.g. the blue/red eye island.

Example 22.2.1.

22.3 Morris, 94 EMCA

Morris, S. (1994). Trade with Heterogeneous Prior Beliefs and Asymmetric Information. Econometrica 62(6), 1327

This paper is an extension of Milgrom and Stokey (1982), which relaxes the concordant belief in MS-1982 paper and find the no-trade theorems under different *trading rules* and *belief structures*.

The generalization is listed in Berk's notes:

C \ T	Ex-Ante	Interim	Ex-Post
Uncons.	condordant	consistent conc.	revelation cons. con.
IC	noisy conc.	noisy cons. conc.	-
Public	public conc.	public cons. conc.	public rev. cons. conc.

The read of this table is: "An initially efficient allocation is T C efficient *iff* beliefs are (C,T)."

- Noisy concordant and public concordant are weaker than concordant. If beliefs are concordant, then they are noisy concordant.

22.4 Kiyotaki, 2011

Kiyotaki, N. (2011). A mechanism design approach to financial frictions. manuscript

This paper mainly discusses the micro structures which may induce financial friction (inefficiency). These structures include private information of income, private technology for storage, limited commitment and their combinations.

Chapter 23

Costly State Verification

Financial frictions is an important yet not well explained phenomena. Generally financial frictions means some possible financial frictions not happening, resulting in fluctuations in real economy. There are several micro foundations for financial frictions, we learned some in Ruilin's course:

- costly state verification
- collateral constraints
- moral hazard
- asymmetric information

The main framework for costly state verification is following. The borrower has some private information on productivity, and lender has to pay some cost to get this information. This will give the borrower some advantage, and make some possible transactions unable to happen. Townsend (1979) gives a micro model describing contract under costly state verification. Bernanke and Gertler (1989) and Carlstrom and Fuerst (1997) are two important works using this to explain business cycles (also called Financial Accelerator).

Add brief introduction for costly state verification literature.

23.1 Townsend, 79 JET

seminal paper in costly state verification, important, to be added.

23.2 Kiyotaki-Moore, 97 JPE

Kiyotaki, N. and J. Moore (1997). Credit cycles. Journal of Political Economy 105(2), 211–248

Another explanation for financial friction is collateral constraint. Important literatures include Kiyotaki and Moore (1997),

Kiyotaki and Moore (1997) has a simple idea. Suppose the borrower needs some collateral (say land) for borrowing. What is the effect of a temporary increase in land productivity? The direct effect is it will increase this period land price. More important is the indirect effect, the borrower now can buy more land since their collateral becomes more valuable. And thus in next period they can buy even more land. This long chain has a large effect on the initial price of the land. Thus financial sector can amplify the fluctuation of real economy.

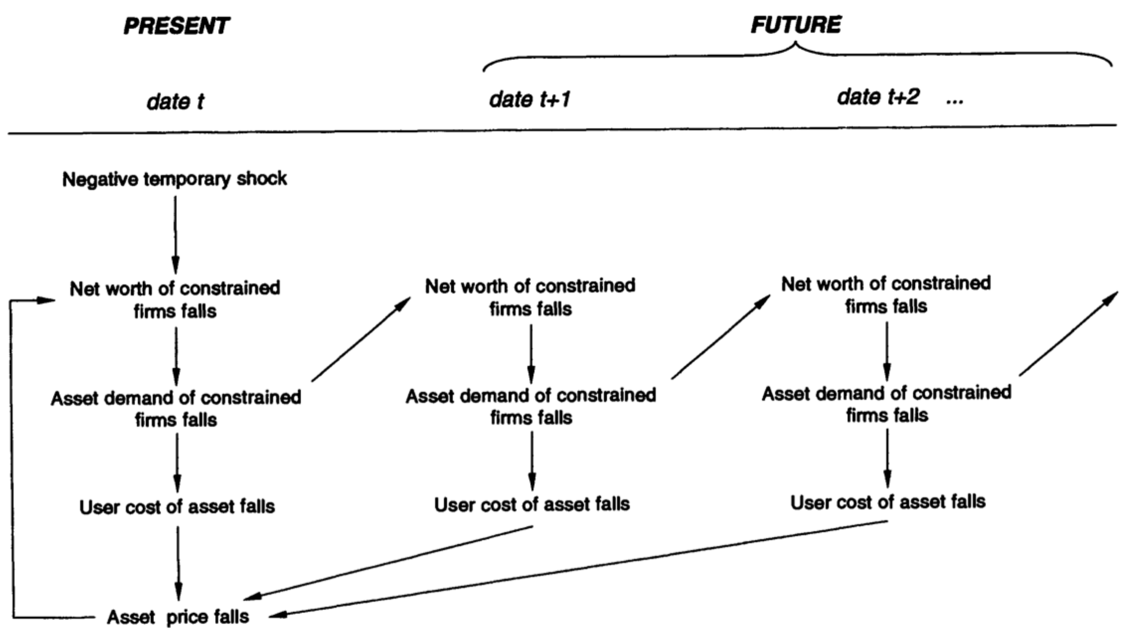


FIG. 1

Skills:

- How they deal with specific-targeted loan (i.e. the loan can only be used to purchase land)

23.3 Gorton and Ordonez, 14 AER

Gorton, G. and G. Ordoñez (2014, February). Collateral Crises. American Economic Review 104(2), 343–378

Motivation: How a small shock can sometimes have a very huge, sudden effect, while at other times the effect of the same size shock is small or nonexistent. Leverage per se is not enough to explain this. This paper explains how credit booms arise, leading to financial fragility where a small shock can lead to huge consequences. Thus, "tail risk" is endogenous. The same reason causes boom and huge recession, and the degree of recession is positively correlated with the boom lasting time.

Mechanism:

1. Firm needs land as collateral to borrow money from households. Land is either good or bad. Neither firm nor household know the quality of land, and checking fee is $\gamma > 0$. It can be shown that, when the probability of a land to be good is too high or too low, it will be optimal (for both firm and household) to not check the quality.

2. Then suppose there are idiosyncratic shocks to land. For each land in each period, there is λ prob to be unchanged, and $1 - \lambda$ to be changed. In case of changing, $\hat{p}$ prob becoming good and $1 - \hat{p}$ becoming bad, independent of initial type. So we can take it as a $\hat{p}$ type, it will become either 0 or 1 after checking. If $\hat{p}$ lies in some range that no checking is appropriate, then uncertainty in the economy will accumulate. Notice such accumulation will lead to a boom.

3. Now think of an aggregate shock. All good type and $\hat{p}$ type have weaker beliefs after such shock, changing to η and $\eta\hat{p}$ respectively. Notice that η and $\eta\hat{p}$ may have very different effect. It may be optimal to not check with η but optimal to check with $\eta\hat{p}$. Thus if an economy has accumulated a large portion of $\hat{p}$, i.e. a long boom, then many of them will choose to check, inducing a huge crisis.

4. Notice that it is not because of any *intrisinc* change in economy. The $\hat{p}$ will change the real portion of good land in the economy only once. What really matters is that once the bad one becomes $\hat{p}$, then even it switches back to bad in following periods, it will still be taken as $\hat{p}$. That means, the $\hat{p}$ type will accumulate in the economy.

Why this paper is important:

- Provide a way to endogenize information in general equilibrium model.

23.4 Williamson, 87 QJE

Williamson, S. D. (1987). Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing. The Quarterly Journal of Economics 102(1), 135–145

Williamson (1987) mainly wants to explain two things:

- Why is debt contract so popular?
  - By MM Theorem, we know that in a frictionless world, debt contract is equivalent to equity contract. But we notice that most fund is raised through debt.
  - Explain by moral hazard. Borrower has private information on outcome. And lender has to pay for checking.
- Why there is credit rationing?
  - If the borrowing demand exceeds lending supply, why doesn't the interest rate increase to clear the market?
  - Because checking is costly, so incentive compatibleness requires interest rate not being too high. They prove that the optimal interest rate is hump shape.

An IC contract $\{S, K(w), R(w)\}$ contains three parts. S is the checking area. $K(w)$ is the transfer to lender when not checking. $R(w)$ is the transfer to lender when checking. Since the contract is IC, so the reported $\hat{w}$ is just actual realized w.

There are three steps to prove the optimal contract is debt contract in this asymmetric information case.

1. Prove $K(w)$ equals a constant x. [easy]
2. Prove the IR constraint of lender will be tight. [easy]
3. Prove $R(w) = w$, i.e. lender will take away everything in checking case. [difficult]

In the third step, the idea is easy. Because if lender takes away everything in checking case, then the checking area can be smaller, thus can save checking cost. The difficulty is: when we lower x (to make checking area A smaller), the non-checking area B becomes larger. The changes are in two different directions and problem becomes subtle. In discussion we use the total differentiation approach to solve this. Important.

Techniques:

how to prove optimal contract format:

See above.

recursive structure in defining EQM:

One difficulty in determining the optimal contract is how to characterize outside options of lender. In equilibrium, When a lender decides whether to accept a contract from a specific borrower (entrepreneur), his outside option should be market interest rate r. Because entrepreneur can't identify lender, and all entrepreneur are homogeneous ex ante, so if the expected return from a specific contract should be larger than mkt interest rate r. Thus the IR of lender is:

$$\int_A R(w)d(F(w)) + \int_B xd(F(w)) - \gamma Pr(A) \geq r$$

r has to be pinned down in the eqm. Because all agents have rational expectation, thus the market clearing r will be the same as their expectation.

The equilibrium is defined as:

-

23.5 Chari et.al, 14 AER

Chari, V. V., A. Shourideh, and A. Zetlin-Jones (2014, December). Reputation and Persistence of Adverse Selection in Secondary Loan Markets. American Economic Review 104(12), 4027–4070

Empirical Facts:

- simultaneous fluctuation between secondary loan market and underlying land market
- persistence of collapse in secondary loan market trade volume

Main Mechanism:

- adverse selection AND reputation
- key assumptions:
  - agent (bank) types persist over time
  - players are non-annonymous (possible to build reputation)
- adverse selection: can account for trade volume fluctuations correlated with underlying collateral value
- reputation:
  - wihout reputation concern, adverse selection will be solved quickly. Because the first stage features a separating equilibrium
  - with high enough reputation concern (patient enough), first stage cannot be fully separate, so adverse selection problem persists

Chapter 24

Euro Crisis

This section is based on a course material of Wang Jue. Basically it discusses the theoretical concerns of Optimal Currency Area (OCA), why Europe adopts it, its success and its failure, and the recent Euro crisis.

Krugman, P. (2012). Revenge of the Optimum Currency Area. In NBER Macroeconomics Annual 2012, Volume 27, pp. 439–448. National Bureau of Economic Research, Inc

Krugman (2012) discusses the crisis of Euro area. Based on OCA theory by Mundell, there are two big things to look at: **labour mobility** and **fiscal integration**. Because areas are likely to suffer local idiosyncratic shocks.

- Without labor mobility, a state has to make wage fall to restore the full employment, this will be much easier with its own to devalue. But if labor mobility is high, then workers can just move to other states.
- Fiscal integration will also help. First because federal transfer is like a insurance, also because the borrowing cost for federal government is much lower.

The Euro crisis faces challenges in both parts.

- The labor mobility is not high enough.
- No fiscal integration. And the asymmetric shock is so severe, making fiscal burden so large that calls government solvency into question. (e.g. in Spain and Greek)
- The bank issue. There is no 'last resort', so the fear of sovereign default undermineds confidence in private banks which hold these bonds. This is called *doom loops*.

- Compare between euro area countries and pegged-to-euro countries. Austria and Finland uses euro, but Finland just pegged to euro. This flexibility makes interest rate much lower during the crisis.

Baldwin, R. E. and F. Giavazzi (2015). The Eurozone Crisis: A Consensus View of the Causes and a Few Possible Remedies. CEPR Press

Chapter 25

Macro Seminar Notes

Cryptocurrency

- By Jesús Fernández-Villaverde and Daniel Sanches 2017 .Use Lagos and Wright (2005) framework.
  - `http://economics.sas.upenn.edu/~jesusfv/Cryptocurrencies.pdf`
  - Not positive on private currencies. First, competitive private money will not provide price stability, either with a profit-maximizing entrepreneurs or through automaton.
  - Even with price stability, value of private money may go zero in infinite equilibriums.
  - Purely private money system does not provide social optimal quantity of money.

Sorkin, I. (2018). Ranking Firms Using Revealed Preference*. The Quarterly Journal of Economics

- Use Google Page Ranking algorithm to rank good firms. 'good firms hire from other good firms and have few workers leave'. Reduce computation burden greatly.
- Previous literature use hedonic approach to rank firms. Two drawbacks:
  - Need to completely model and measure all relevant elements
  - Do not consider the utility dispersion case. [not understand]

Worth reading.

Part VI

Methodology

Chapter 26

Micro Foundations of Macroeconomics

Reductionism, and Micro Fundations of Economics:

Economists are fascinated in forming 'micro fundations' for almost all models. But do we really need this, does this effort pay?

As pointed in famous paper Anderson (1972): '*The ability to reduce everything to simple fundamental laws doesn't imply the ability to start from those laws and reconstruct the universe... At each level of complexity entirely new properties appear... There is no useful parallel to be drawn between the way in which complexity appears in simplest cases of many modies and chemistry and culture ones.*' The aggregated economy is so complicated that may not be reduced to individual decision problems. [1]

Another problem worth thinking is the relationship between Sonnethein-Debreu theorem and macro economics models.

Policy Making without Macroeconomic Theory: [2]

Kocherlakota (2018) points out that policy makers does not need to know economic theory, albeit the players are in the framework of theory. The key points of the theory are:

- Private sectors are forward looking, but government is myopic

[1] Zhao Dingxing has an article discussing this problem.
See `https://www.evernote.com/l/ATUOIdQb8hBM5r3QQv8cTK52Amq4-qlwb0E`
[2] Click for link to slides

- Gov suffers private taste shock, which will only affect the gov utility
- Within an equilibrium, government can pick the optimal policy through simple regression over past actions and stages

Chapter 27

Decision Theory as a Foundation

Meaning of Decision Theory:

(This part is heavily influenced by Henrique's class.)

Why we need to do decision theory? Take an *inverse engineering* perspective. Suppose we have a model, which uses utility function to describe people's decison choices. That is, $u : X \to \mathbb{R}$. And we define $C(B) = \arg\min_{x \in B} u(x) = \{x \in B | u(x) \geq u(y), \forall y \in B\}$. We can map such functions to a *equivalent class*, and can further map it to a *preference relation* (a subset of relations).

The problem is such mapping is not surjective. Every utility function has a corresponding equivalence and a prefence relation, but there is some preference relation that cannot be represented by a utility function. Thus we need to put some restrictions over preference relation.

The first relation is that the cardinality of indifference curves cannot be larger than $\mathbb{R}$. This is easy to understand.

But this restriction is not enough. Think of lexicographic preference, every point in the two-dimension space is a indifference class. The cardinality of which is R^2, equivalent to cardinality of $\mathbb{R}$. But we know that such preference do not have a corresponding utility function. We can further restrict preference to *order separable* relations, then by Debreu Thm, we can show the *iff* between utility and preference relations.

Definition 27.0.1 (Order Separable). A preference relation defined over X is order separable if $\exists Y \subseteq X$ countable s.t. $\forall x_1, x_3 \subseteq X$, we have $x_1 \succ x_2 \Rightarrow \exists y \in Y$ s.t. $x_1 \succ y \succ x_2$.

Important here is thaat by assuming the range of utility function is $\mathbb{R}$ (such a weak condition), we actually put some restrictions over preference.

Calibration:

Why use calibration instead of estimation?

- Because we live in a 'tenth-best' world, and our model only catches a small part of the world. Use calibration we can fit the moments that our model fail to catch, in this sense to be more accurate.

May use calibration parameters as an intuitive check of whether the code we wrote is correct.

Research Reproducibility

Maniadis and Tufano (2017) claims two aspects of reproducibility problem. First, it can be seen as an 'evidence game', where researchers have some private information and evaluators have to design a incentive compatible mechanism for publishing and granting. Second, empirical methods from meta-research may help. (I don't understand the second part.)

Maniadis et al. (2017) in the same issue provides more detailed guide on what will affect the credibility of scientific research. It raises a concept post-study probability,

$$PSP = \frac{(1-\beta)\pi}{(1-\beta)\pi + \alpha(1-\pi)}$$

where α is significance level, $1-\beta$ is power, and π is the probability that a research question is actually true. π can be think of as *prior* given by theory. This is the fraction of 'truly true and reported true' divided by 'all reported true'. Easy to see, when π is low, the PSP can be very low even with low α. This indicates that with those fields where theory can provide little guide, we have to be very skeptical to the 'rejection' outcomes. They are very likely to be 'false positive'.

Spiess (2017) has a fantastic discussion of this from the perspective of econometrics and mechanism design.

Lucas Critique:

Hoover (1994) views Lucas critique as pointing the causal inference and identification problem in econometrics.

Part VII

Mathematics

Chapter 28

General (mathematical) Skills

28.1 Log-Linearization

Motivation

Why we need log linearizations? Because for most nonlinear discrete dynamic programming problems, we fail to find a closed solution. Thus we have to use numerical method or to find a approximation. By using log-linearization, we transform the nonlinear problem to a linear problem (around the steady state), and we know how to solve the linear diffrence equations.

Another advantege of log-linearization is, the new variables are in forms of $\frac{x-x^*}{x^*}$, which is interpretable.

28.1.1 How to do Log-Linearization

Log linearization is a common method to approximate non-linear function using Taylor expansion.

When we do linearization in macroeconomics, one key is to find **steady state** of the model. Then we do linearization aroud the steady state.

The basic idea is, for functions like:

$$f(x) = \frac{g(x)}{h(x)}$$

take log in both sides:

$$ln(f(x)) = ln(g(x)) - ln(h(x))$$

By Taylor expansion, for a smooth function f(x) we have:

$$f(x) = f(x^*) + \frac{f'(x^*)}{1!}(x - x^*) + o(1)$$

Thus

$$lnf(x) \approx lnf(x^*) + \frac{f'(x^*)}{f(x^*)}(x - x^*)$$

The above follows from the fact that $\frac{dln(f(x))}{dx} = \frac{f'(x)}{f(x)}$. Thus,

$$lnf(x) - lnf(x^*) = \frac{f'(x^*)}{f(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*}$$

And the equation becomes:

$$lnf(x^*) + \frac{f'(x^*)}{f(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*} = lng(x^*) - lnh(x^*)\frac{g'(x^*)}{g(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*} - \frac{h'(x^*)}{h(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*}$$

$$\frac{f'(x^*)}{f(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*} = \frac{g'(x^*)}{g(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*} - \frac{h'(x^*)}{h(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*}$$

Dedine $\hat{x} = \frac{x - x^*}{x^*}$, and we are done.

$$\frac{f'(x^*)}{f(x^*)}x^* \cdot \hat{x} = \frac{g'(x^*)}{g(x^*)}x^* \cdot \hat{x} - \frac{h'(x^*)}{h(x^*)}x^* \cdot \hat{x}$$

Notice, when we use $\hat{x}$ in place of $x - x^*$in the approximation, don't forget to times x^* !!

28.1.2 log-linearization when there is expectation (An Example)

It may be annoying to do log-linearization when there is expectation, because in general, log and expectation cannot switch. But if we don't switch, we cannot do the log-linearization as previous. For example, see Adda&Cooper(2003) CH5, page 115.

Suppose we have the following Euler equation:

$$u'(c) = \beta E_{A'|A} u'(c)[Af'(k') + (1-\delta)]$$

where $c = Af(k) + (1-\delta)k - k'$.

These two expressions, along with the evolution of A, defines a system of equations.

In this system, actually we do not have to take log, but directly do Taylor expansion at c^*, x^* and k^*. We get:

$$\begin{aligned} u'(c^*) + u''(c^*)c^*\hat{c}_t = \beta E_{A'|A}[u'(c^*)(\bar{A}f'(k^*) + (1-\delta)) + (\bar{A}f'(k^*) + (1-\delta))u''(c^*)c^*\hat{c}_{t+1} \\ + u'(c^*)f'(k^*)\bar{A}\hat{A}_{t+1} + u'(c^*)\bar{A}f''(k^*)k^*\hat{k}_{t+1}] \end{aligned} \quad (28.1)$$

The first term in LHS and the first term in RHS cancel each other. And we divide both sides by $u'(c^*)$, and we will get:

$$\frac{u''(c^*)c^*}{u'(c^*)}\hat{c}_t = \beta E_{A|A'}[\frac{u''(c^*)c^*}{u'(c^*)}\hat{c}_{t+1}\frac{1}{\beta} + f'(k^*)\bar{A}\hat{A}t + 1 + f''(k^*)k^*\hat{k}_{t+1}] \quad (28.2)$$

Notice, we derive this based on the fact $\frac{1}{\beta} = \bar{A}f'(k^*) + (1-\delta)$. Because by FOC

$$u'(c) = \beta E_{A'|A} V'(k')$$

To derive $V'(k')$, we first derive $V'(k)$, by Envelop THM we get:

$$V'(k) = u'(c)[Af'(k) + (1-\delta)]$$

Substitute this into FOC we get:

$$u'(c) = \beta E_{A|A'}[u'(c')(A'f'(k') + (1-\delta))]$$

In this approach we fix A at the mean $\bar{A}$, and thus the steady state satisfy:

$$1 = \beta[\bar{A}f'(k^*) + (1-\delta)]$$

Here, the introduce of $\bar{A}$ is to describe the steady state in existence of shock. We can see $\bar{A}$ as the "long term" growth of the economy, and the we investigate what's the optimal steady k^* under this long-term growth rate.

But I still don't know why we can cancel the expectation in the RHS.

The equation 28.2 is the linear function we want.

Some Comments on the above method:

Although it seems that the above method uses no log-linearization, actually it does!

(to be added later)

28.2 Supermodularity

Reference:

- Sobel's Notes: `http://econweb.ucsd.edu/~jsobel/205f10/notes12.pdf`
- A more detailed note: `https://pages.wustl.edu/files/pages/imce/nachbar/monotonecomparativestatics.pdf`
- Topkis, D. (1998). Supermodularity and Complementarity, 1998. Princeton University Press

28.3 Miscellanea

28.3.1 Elasticity

Example 28.3.1. Suppose $Q = \sum q_i$, $p = Q^{1/\alpha}$ and define $s_i = \frac{q_i}{Q}$. In equilibrium, we have $s_i = s_i^*$. Then in equilibrium, we have $\frac{\partial p}{\partial q_i} = s_i^*/\alpha$ instead of $1/\alpha$.

Why? In first look, Q is linear in q_i, then the derivative $\partial p/\partial Q$ should be the same as $\partial p/\partial q_i$. The point is that Q is just a aggregated notation. In equilibrium, small change in q_i will induce small change in all $q_{j\neq i}$, thus the aggregate change in Q will be proportion of change in q_i.

Chapter 29

Probability

Theorem 29.0.1 (Linear Combination of Dependent Normal RV). *Suppose* $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \Sigma_{X,Y}\right]$, *then* $AX + BY + C \sim \mathcal{N}\left[\begin{pmatrix} A & B \end{pmatrix}\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} + C, \begin{pmatrix} A & B \end{pmatrix}\Sigma_{X,Y}\begin{pmatrix} A^T \\ B^T \end{pmatrix}\right]$.

Theorem 29.0.2 (Kolmogorov's zero–one law). *Let* (Ω, F, P) *be a probability space, and* F_n *be a series of independent* σ*-algebras contained in* F*. Let*

$$G_n = \sigma\left(\cup_{k=n}^{\infty} F_k\right)$$

Then we define tail event F *to be any event such that:*

$$F \in \cap_{n=1}^{\infty} G_n$$

Kolmogorov zero-one law claims that $Pr(F) = 0$ *or* $Pr(F) = 1$*.*

Intuitively, if the occurrence of a event F can be determined when any finite initial segment is removed, then it either happens or not with probability 1.

29.1 Distributions

- Pareto Distribution
- T1EV (Type 1 Extreme Value)

29.2 Bayesian Statistics

29.2.1 Bayesian Updating

Theorem 29.2.1 (Bayesian Update in Odds). *For hypothesis H, data d and prior odds $O(H)$. We have posterior odds $O(H|d) = O(H) \cdot \frac{Pr(d|H)}{Pr(d|H^c)}$, where the second term is called Bayes Factor.*

Proof:

$$\begin{aligned} O(H|d) &= \frac{Pr(H|d)}{Pr(H^c|d)} \\ &= \frac{Pr(d|H)Pr(H)}{Pr(d|H^c)Pr(H^c)} \\ &= \frac{Pr(d|H)}{Pr(d|H^c)} \cdot O(H) \end{aligned}$$

□

Chapter 30

Numerical Methods

Computation laguage:

- `http://economics.sas.upenn.edu/~jesusfv/comparison_languages.pdf`
  - discusses efficiency of different languages
  - hybrid language can significantly increase efficiency

Matlab Speed Up:

- Microeconometrics and Matlab: an Introduction, Ch-10
- vectorize and preallocate values
  - Cooper's growth model uses clever vectorizing, avoid one loop
- profiling code: profile on, main, profile off
- parallel computing

30.1 Numerical Maximization

Newton-Raphson Method:

Idea: use quadratic function to approaximating real function at current guess, then find the maximization point of this quadratic function as next guess.

$$\beta_{t+1} = \beta_t + \lambda(-H_t^{-1})g_t$$

where H_t is the Hessian of the function to be evaluated at β_t, and g_t is the Jacobian at β_t, λ is the step size, generally may take 1.

BHHH Method:

Idea: almost the same as Newton-Raphson, but use the outer product to approximate Hessian.

$$\beta_{t+1} = \beta_t + \lambda(-B_t^{-1})g_t$$

where $B_t = \frac{1}{N}\sum_{n=1}^{N} s_n(\beta_t)s_n(\beta_t)'$ and $s_n(\beta_t) = \partial P_n(\beta)/\partial\beta$ evaluated at β_t, and $g_t = \frac{1}{N}\sum_{n=1}^{N} s_n(\beta_t)$.

Key is, when evaluated at maximum value, s_n should be zero. Thus B_t evaluated at maximum is the variance of s_n in the sample, which is just the Hessian of the sample likelihood function. This is just *information equality*.

pros: BHHH is generally faster than Newton-Raphson near the maximum. Because s_n have to be calculated anyway, but in NR procedure we have to calculate the second derivatives as well. But seems this is not the case, in experiments BHHH always slower than Newton-Raphson.

cons: BHHH is generally slower when β_t is far from true value. Because at these points B_t is not a good approximate of H_t.

- Root Solving
- Integration
  - Gaussian quadrature
  - Sparse grid
  - Monte Carlo integration
    - antithetic accelerating
    - importantce sampling
  - Quasi Monte Carlo

Chapter 31

Miscellany

Presentation:

- Always *motivate* very clearly, both from economics facts and from literature.
  - In macro, a graph generally helps
- Make clear propositions so that others won't misunderstand.
  - Bad: analyze the competitive pattern of inter-brand and intra-brand

    Good: Analyze how the per store profit is affected by own-brand store density and competitors' store density
- If possible, always use a simple model at the beginning
  - To pin down the concepts/definitions
  - To show the mechanism intuitively
- Understand the result both mathematically and intuitively
- Visualization:
  - If only the scale of number is important, use a *plot* instead of table full of numbers
- When writing on the board, write horizontally!

Writings:

- Good writing samples: Berry (1992), Holmes (2011)

– Holmes (2011): a good example of illustrating clear empirical results

Bibliography

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. Econometrica 83(6), 2411–2451.

Allen, J., R. Clark, and J.-F. Houde (2014, October). The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry. American Economic Review 104(10), 3365–3396.

Anderson, P. W. (1972). More is different. Science 177(4047), 393–396.

Athey, S., J. Levin, and E. Seira (2011, February). Comparing open and Sealed Bid Auctions: Evidence from Timber Auctions. The Quarterly Journal of Economics 126(1), 207–257.

Bajari, P., C. L. Benkard, and J. Levin (2007, September). Estimating Dynamic Models of Imperfect Competition. Econometrica 75(5), 1331–1370.

Baldwin, R. E. and F. Giavazzi (2015). The Eurozone Crisis: A Consensus View of the Causes and a Few Possible Remedies. CEPR Press.

Baye, M. R., B. De los Santos, and M. R. Wildenbeest (2013). The Evolution of Product Search. The Journal of Law, Economics Policy 9(2), 1–21.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2014, December). Optimal Allocation with Costly Verification. American Economic Review 104(12), 3779–3813.

Berger, J., A. T. Sorensen, and S. J. Rasmussen (2010, September). Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. Marketing Science 29(5), 815–827.

Bernanke, B. and M. Gertler (1989, March). Agency Costs, Net Worth, and Business Fluctuations. American Economic Review 79(1), 14–31.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. Econometrica 63(4), 841–890.

Berry, S. T. (1992). Estimation of a Model of Entry in the Airline Industry. Econometrica 60(4), 889.

Berry, S. T. (1994, July). Estimating Discrete-Choice Models of Product Differentiation. 25(2), 242–262.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2010, December). The financial reporting environment Review of the recent literature. Journal of Accounting and Economics 50(2-3), 296–343.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. Journal of Political Economy 100(5), 992–1026.

Bresnahan, T. F. and P. C. Reiss (1990). Entry in Monopoly Markets. The Review of Economic Studies 57(4), 531.

Butters, G. R. (1977). Equilibrium Distributions of Sales and Advertising Prices. The Review of Economic Studies 44(3), 465–491.

Cardon, J. H. and I. Hendel (2001). Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey. The RAND Journal of Economics 32(3), 408.

Carlstrom, C. T. and T. S. Fuerst (1997, December). Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis. American Economic Review 87(5), 893–910.

Chari, V. V., A. Shourideh, and A. Zetlin-Jones (2014, December). Reputation and Persistence of Adverse Selection in Secondary Loan Markets. American Economic Review 104(12), 4027–4070.

Chatterjee, S. (2018). Market Power and Spatial Competition in Rural India.

Chiappori, P. A. and B. Salanie (2000, February). Testing for Asymmetric Information in Insurance Markets. Journal of Political Economy 108(1), 56–78.

Chiong, K. X., A. Galichon, and M. Shum (2016, April). Duality in dynamic discrete-choice models. Quantitative Economics 7(1), 83–115.

Ciliberto, F. and E. Tamer (2009). Market Structure and Multiple Equilibria in Airline Markets. Econometrica 77(6), 1791–1828.

Copeland, T. E. and D. Galai (1983). Information effects on the bid-ask spread. the Journal of Finance 38(5), 1457–1469.

Cramton, P. and J. A. Schwartz (2000, May). Collusive Bidding: Lessons from the FCC Spectrum Auctions. Journal of Regulatory Economics 17(3), 229–252.

d'Aspremont, C., J. J. Gabszewicz, and J. F. Thisse (1979). On Hotelling's "Stability in Competition". Econometrica 47(5), 1145–1150.

Davis, P. (2006, September). Estimation of quantity games in the presence of indivisibilities and heterogeneous firms. Journal of Econometrics 134(1), 187–214.

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2017). Revealed price preference: Theory and stochastic testing.

Dye, R. A. (1985). Disclosure of Nonproprietary Information. Journal of accounting research 23(1), 123–145.

Einav, L., A. Finkelstein, and M. R. Cullen (2010). Estimating Welfare In Insurance Markets Using Variation in Prices. Quarterly Journal of Economics 125(3), 877–921.

Einav, L. and J. Levin (2010, May). Empirical Industrial Organization: A Progress Report. Journal of Economic Perspectives 24(2), 145–162.

Ellison, G. (1994). Theories of Cartel Stability and the Joint Executive Committee. The RAND Journal of Economics 25(1), 37–57.

Erdem, T. and M. P. Keane (1996). Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. Marketing Science 15(1), 1–20.

Fainmesser, I. P. and A. Galeotti (2018). The Market for Influence. SSRN Electronic Journal, 1–40.

Fudenberg, D. and D. K. Levine (1993). Self-Confirming Equilibrium. Econometrica 61(3), 523–545.

Gale, D. and L. S. Shapley (1962, January). College Admissions and the Stability of Marriage. The American Mathematical Monthly 69(1), 9–15.

Gandhi, A., S. Navarro, and D. Rivers (2017). On the Identification of Gross Output Production Functions.

Geanakoplos, J. D. and H. M. Polemarchakis (1982). We can't disagree forever. Journal of Economic Theory 28(1), 192–200.

Gebru, T., J. Krause, Y. Wang, D. Chen, J. Deng, E. L. Aiden, and L. Fei-Fei (2017, February). Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US. arxiv.org.

Genesove, D. and W. P. Mullin (2001, June). Rules, Communication, and Collusion: Narrative Evidence from the Sugar Institute Case. American Economic Review 91(3), 379–398.

Gentzkow, M. and E. Kamenica (2017, January). Competition in Persuasion. The Review of Economic Studies 84(1), 300–322.

Gentzkow, M., J. Shapiro, and M. Taddy (2017, May). Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech.

Gentzkow, M., M. Taddy, and B. T. Kelly (2017, February). Text as Data. pp. 1–53.

Gillen, B. J., S. Montero, H. R. Moon, and M. Shum (2018). BLP-Lasso for Aggregate Discrete Choice Models of Elections with Rich Demographic Covariates.

Glazer, J. and A. Rubinstein (2004). On Optimal Rules of Persuasion. Econometrica 72(6), 1715–1736.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of financial economics 14(1), 71–100.

Gorton, G. and G. Ordoñez (2014, February). Collateral Crises. American Economic Review 104(2), 343–378.

Graham, D. A. and R. C. Marshall (1987). Collusive Bidder Behavior at Single-Object Second-Price and English Auctions. Journal of Political Economy 95(6), 1217–1239.

Green, E. J. and R. H. Porter (1984). Noncooperative Collusion under Imperfect Price Information. Econometrica 52(1), 87–100.

Green, J. R. and J.-J. Laffont (1986). Partially Verifiable Information and Mechanism Design. The Review of Economic Studies 53(3), 447.

Grieco, P. L. E. (2014, June). Discrete games with flexible information structures: an application to local grocery markets. The RAND Journal of Economics 45(2), 303–340.

Grimmer, J. and B. M. Stewart (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political Analysis 21(3), 267–297.

Grossman, S. J. (1981, December). The Informational Role of Warranties and Private Disclosure about Product Quality. The Journal of Law & Economics 24(3), 461–483.

Hadfield, G. K. and B. R. Weingast (2012, November). What Is Law? A Coordination Model of the Characteristics of Legal Order. Journal of Legal Analysis 4(2), 471–514.

Hart, S., I. Kremer, and M. Perry (2017, March). Evidence Games: Truth and Commitment. American Economic Review 107(3), 690–713.

Hausman, J., G. Leonard, and J. D. Zona (1994, April). Competitive Analysis with Differenciated Products. Annales d'Économie et de Statistique (34), 159–180.

Holmes, T. J. (2011). The Diffusion of Wal-Mart and Economies of Density. Econometrica 79(1), 253–302.

Hoover, K. D. (1994). Econometrics as observation: the Lucas critique and the nature of econometric inference. Journal of Economic Methodology 1(1), 65–80.

Horn, H. and A. Wolinsky (1988). Bilateral Monopolies and Incentives for Merger. The RAND Journal of Economics 19(3), 408.

Hortacsu, A. and C. Syverson (2004, May). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. The Quarterly Journal of Economics 119(2), 403–456.

Hotz, V. J. and R. A. Miller (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. The Review of Economic Studies 60(3), 497.

Iaryczower, M., X. Shi, and M. Shum (2018). Can Words Get in the Way? The Effect of Deliberation in Collective Decision-Making. Journal of Political Economy.

Jia, P. (2008). What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry. Econometrica 76(6), 1263–1316.

Kamenica, E. and M. Gentzkow (2011, October). Bayesian Persuasion. American Economic Review 101(6), 2590–2615.

Keane, M. P. and K. I. Wolpin (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. The Review of Economics and Statistics 76(4), 648–672.

Kiyotaki, N. (2011). A mechanism design approach to financial frictions. manuscript.

Kiyotaki, N. and J. Moore (1997). Credit cycles. Journal of Political Economy 105(2), 211–248.

Kremer, I., Y. Mansour, and M. Perry (2014). Implementing the "Wisdom of the Crowd". Journal of Political Economy 122(5), 988–1012.

Krishna, V. and R. W. Rosenthal (1996). Simultaneous Auctions with Synergies. Games and Economic Behavior 17(1), 1–31.

Krugman, P. (2012). Revenge of the Optimum Currency Area. In NBER Macroeconomics Annual 2012, Volume 27, pp. 439–448. National Bureau of Economic Research, Inc.

Levinsohn, J. and A. Petrin (2003, April). Estimating Production Functions Using Inputs to Control for Unobservables. The Review of Economic Studies 70(2), 317–341.

Lewbel, A. (2016). The Identification Zoo-Meanings of Identification in Econometrics. Forthcoming on Journal of Economic Literature.

Low, H. and C. Meghir (2017, May). The Use of Structural Models in Econometrics. Journal of Economic Perspectives 31(2), 33–58.

Maniadis, Z. and F. Tufano (2017, October). The Research Reproducibility Crisis and Economics of Science. The Economic Journal 127(605), F200–F208.

Maniadis, Z., F. Tufano, and J. A. List (2017, October). To Replicate or Not To Replicate? Exploring Reproducibility in Economics through the Lens of a Model and a Pilot Study. The Economic Journal 127(605), F209–F235.

Marshall, R. C., M. J. Meurer, J. F. Richard, and W. Stromquist (1994). Numerical Analysis of Asymmetric First Price Auctions. Games and Economic Behavior 7(2), 193–220.

McMillan, J. (1994). Selling Spectrum Rights. Journal of Economic Perspectives 8(3), 145–162.

Milgrom, P. and N. Stokey (1982). Information, trade and common knowledge. Journal of Economic Theory 26(1), 17–27.

Milgrom, P. R. (1981, October). Good News and Bad News: Representation Theorems and Applications. The Bell Journal of Economics 12(2), 380–391.

Morris, S. (1994). Trade with Heterogeneous Prior Beliefs and Asymmetric Information. Econometrica 62(6), 1327.

Morris, S. (1995, October). The Common Prior Assumption in Economic Theory. Economics and Philosophy 11(2), 227–253.

Nevo, A. (2000, December). A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand. Journal of Economics & Management Strategy 9(4), 513–548.

Ni, J., S. A. Neslin, and B. Sun (2012, November). Database Submission—The ISMS Durable Goods Data Sets. Marketing Science 31(6), 1008–1013.

O'hara, M. (1998). Market microstructure theory, Volume 108. Blackwell Cambridge, MA.

Olley, G. S. and A. Pakes (1996, November). The Dynamics of Productivity in the Telecommunications Equipment Industry. Econometrica 64(6), 1263–1297.

Peters, M. and A. Siow (2002). Competing Premarital Investments. Journal of Political Economy 110(3), 592–608.

Porter, R. H. (1983). A Study of Cartel Stability: The Joint Executive Committee, 1880-1886. The Bell Journal of Economics 14(2), 301–314.

Quan, T. W. and K. R. Williams (2017). Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail. SSRN Electronic Journal, 1–74.

Quigley, D. and A. Walther (2017, May). Inside and Outside Information.

Rey, P. and J. Stiglitz (1995, October). The Role of Exclusive Territories in Producers' Competition. The RAND Journal of Economics 26(3), 431–451.

Rotemberg, J. J. and G. Saloner (1986). A Supergame-Theoretic Model of Price Wars during Booms. American Economic Review 76(3), 390–407.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. Econometrica 55(5), 999.

Schmalensee, D. S. E. and Richard (2013, October). The Antitrust Analysis of Multi-Sided Platform Businesses. manuscript, 1–45.

Seim, K. (2006, September). An empirical model of firm entry with endogenous product-type choices. The RAND Journal of Economics 37(3), 619–640.

Shum, M. (2016). Econometric Models for Industrial Organization. World Scientific.

Smith, L. and P. Sørensen (2000). Pathological outcomes of observational learning. Econometrica.

Sorensen, A. T. (2000). Equilibrium Price Dispersion in Retail Markets for Prescription Drugs. Journal of Political Economy 108(4), 833–850.

Sorkin, I. (2018). Ranking Firms Using Revealed Preference*. The Quarterly Journal of Economics.

Spiess, J. (2017, December). Optimal Estimation when Researcher and Social Preferences are Misaligned. pp. 1–76.

Tamer, E. (2003). Incomplete Simultaneous Discrete Response Model with Multiple Equilibria. Review of Economic Studies 70(1), 147–165.

Topkis, D. (1998). Supermodularity and Complementarity, 1998. Princeton University Press.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. Journal of Economic Theory 21(2), 265–293.

van Damme, E. (1987). Stability and perfection of Nash equilibria, Volume 339. Springer.

Weber, R. J. (1997). Making More from Less: Strategic Demand Reduction in the FCC Spectrum Auctions. Journal of Economics & Management Strategy 6(3), 529–548.

Williamson, S. D. (1987). Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing. The Quarterly Journal of Economics 102(1), 135–145.